

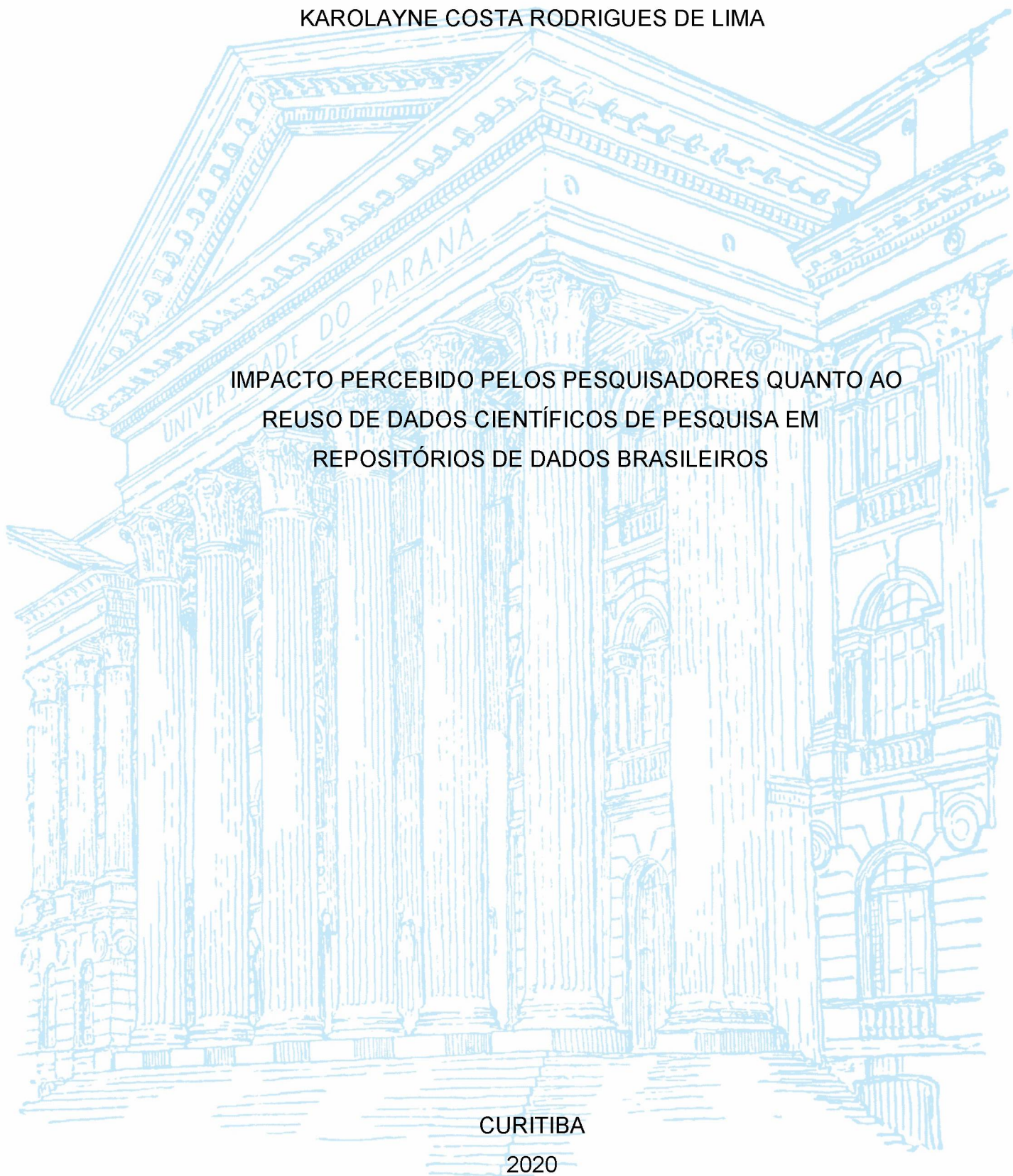
UNIVERSIDADE FEDERAL DO PARANÁ

KAROLAYNE COSTA RODRIGUES DE LIMA

IMPACTO PERCEBIDO PELOS PESQUISADORES QUANTO AO
REUSO DE DADOS CIENTÍFICOS DE PESQUISA EM
REPOSITÓRIOS DE DADOS BRASILEIROS

CURITIBA

2020



KAROLAYNE COSTA RODRIGUES DE LIMA

IMPACTO PERCEBIDO PELOS PESQUISADORES QUANTO AO
REUSO DE DADOS CIENTÍFICOS DE PESQUISA EM
REPOSITÓRIOS DE DADOS BRASILEIROS

Dissertação apresentada ao curso de Pós-Graduação em Gestão da Informação, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Gestão da Informação.

Orientador: Prof. Dr. Ricardo Mendes Junior.

CURITIBA

2020

Catalogação na publicação
Sistema de Bibliotecas UFPR
Biblioteca de Artes, Comunicação e Design/Batel
(Elaborado por: Karolayne Costa Rodrigues de Lima CRB 9/1638)

Lima, Karolayne Costa Rodrigues de
Impacto percebido pelos pesquisadores quanto ao reuso de dados científicos de pesquisa em repositório de dados brasileiros. / Karolayne Costa Rodrigues de Lima. – Curitiba, 2020.

1 arquivo digital (178 p.) : PDF ; 3.98MB

Dados científicos: <http://dx.doi.org/10.5380/bdc/54>

Disponível em World Wide Web.

Orientador: Prof. Dr. Ricardo Mendes Junior.

Dissertação (mestrado em Gestão da Informação) – Universidade Federal do Paraná, Setor de Ciências Sociais Aplicadas, Programa de Pós-Graduação em Gestão da Informação.

1. Dados científicos de pesquisa. 2. Reuso de dados científicos – Impacto. 3. Compartilhamento de dados. 4. Repositório de dados. I. Título.

CDD 025.0427

ATA Nº072020

**ATA DE SESSÃO PÚBLICA DE DEFESA DE MESTRADO PARA A OBTENÇÃO DO
GRAU DE MESTRE EM GESTÃO DA INFORMAÇÃO**

No dia dezesseis de junho de dois mil e vinte às 14:00 horas, na sala <https://conferenciaweb.rnp.br/spaces/ppggi-ufpr>, Sala da Comunidade do PPGGI na Webconf da RNP, foram instaladas as atividades pertinentes ao rito de defesa de dissertação da mestranda **KAROLAYNE COSTA RODRIGUES DE LIMA**, intitulada: **IMPACTO PERCEBIDO PELOS PESQUISADORES QUANTO AO REUSO DE DADOS CIENTÍFICOS DE PESQUISA EM REPOSITÓRIOS DE DADOS BRASILEIROS**, sob orientação do Prof. Dr. RICARDO MENDES JUNIOR. A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO DA INFORMAÇÃO da Universidade Federal do Paraná, foi constituída pelos seguintes Membros: RICARDO MENDES JUNIOR (UNIVERSIDADE FEDERAL DO PARANÁ), TAIANE RITTA COELHO (UNIVERSIDADE FEDERAL DO PARANÁ), MARCOS SFAIR SUNYE (UNIVERSIDADE FEDERAL DO PARANÁ). A presidência iniciou os ritos definidos pelo Colegiado do Programa e, após exarados os pareceres dos membros do comitê examinador e da respectiva contra argumentação, ocorreu a leitura do parecer final da banca examinadora, que decidiu pela APROVAÇÃO. Este resultado deverá ser homologado pelo Colegiado do programa, mediante o atendimento de todas as indicações e correções solicitadas pela banca dentro dos prazos regimentais definidos pelo programa. A outorga de título de mestre está condicionada ao atendimento de todos os requisitos e prazos determinados no regimento do Programa de Pós-Graduação. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, RICARDO MENDES JUNIOR, lavrei a presente ata, que vai assinada por mim e pelos demais membros da Comissão Examinadora.

CURITIBA, 16 de Junho de 2020.

Assinatura Eletrônica

19/06/2020 09:35:16.0

RICARDO MENDES JUNIOR

Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

17/06/2020 16:18:37.0

TAIANE RITTA COELHO

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

03/11/2020 13:56:00.0

MARCOS SFAIR SUNYE

Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO DA INFORMAÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **KAROLAYNE COSTA RODRIGUES DE LIMA** intitulada: **IMPACTO PERCEBIDO PELOS PESQUISADORES QUANTO AO REUSO DE DADOS CIENTÍFICOS DE PESQUISA EM REPOSITÓRIOS DE DADOS BRASILEIROS**, sob orientação do Prof. Dr. RICARDO MENDES JUNIOR, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 16 de Junho de 2020.

Assinatura Eletrônica

19/06/2020 09:35:16.0

RICARDO MENDES JUNIOR

Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

17/06/2020 16:18:37.0

TAIANE RITTA COELHO

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

03/11/2020 13:56:00.0

MARCOS SFAIR SUNYE

Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

À Deus e à minha família!

AGRADECIMENTOS

Agradecer é a forma direta de reconhecer e render homenagem àqueles que nos auxiliaram em alguma uma jornada, e neste caso, na jornada chamada dissertação. Assim, eu expresso toda a minha gratidão, carinho e reconhecimento nos agradecimentos abaixo.

Agradeço imensamente à minha família por todo apoio, acolhimento e incentivo, mesmo à distância. Em especial aos meus pais, Jaciara Lima e Sergio Lima, e meu irmão Davi Lima. O incentivo, amor e carinho de vocês fizeram-me forte para acreditar nessa jornada e poder concluí-la.

Um agradecimento muito especial à Franciane Almeida por ter acompanhado, ajudado e acreditado nessa jornada nos momentos em que eu duvidava. Obrigada por compreender os momentos de ausência, e especialmente obrigada por todos os cafés e jantares maravilhosos.

Expresso toda a minha gratidão ao meu orientador, prof. Dr. Ricardo Mendes Junior, por todo apoio, suporte e paciência nessa jornada. Muito obrigada!

Agradeço aos membros da banca examinadora: professora Dra. Taiane Ritta Coelho e professor Dr. Marcos Sfair Sunyê por aceitarem o convite neste momento tão delicado do país e agradeço também por suas contribuições e apontamentos tão acertados e precisos.

Com muito carinho, agradeço aos meus queridos amigos: Aluíso Gonçalves, Elisabete Ferreira, Eglem Veroneze, Fabiane Führ, Fernando Moreira, Lucas Gonçalves, Marilda Carcereri, Paula Carina de Araújo, Patrícia Ferreira, Rogéria Veroneze, Suzana Zulpo e Thalita Marinho, que presencialmente ou em espaços virtuais apoiaram-me nos momentos de incerteza, lágrimas, viagens, boas risadas e que foram generosos ao dividir os momentos de alegria.

Agradeço aos meus colegas no mestrado, especialmente Ícaro Vieira, Igor Martins, Jenifer Grieger, Pedro Campos e Rogério Yamauti por trazerem leveza nos momentos de dúvida. Minha jornada não teria sido a mesma sem a companhia de vocês.

Agradeço com muito carinho à Simone Batista, secretária do PPGGI/UFPR por todo auxílio, carinho e amizade. Estendo o agradecimento a todos os professores do PPGGI/UFPR cuja gentileza em compartilhar o conhecimento foi fundamental para a conclusão desta pesquisa.

Agradeço às ex-diretoras do Sistema de Bibliotecas da UFPR, Ligia Setenareski e Tania Baggio, por todo apoio, incentivo e oportunidades na realização deste trabalho. Muito obrigada!

Agradeço também e com muito carinho às minhas colegas na Biblioteca de Artes, Comunicação e Design, Cristiane Silva, Valkyria Karpinski e Élen Oliveira pelo companheirismo, incentivo, paciência. Obrigada pelo café maravilhoso (e precioso) e as mimosas comidas sob o sol.

Obrigada à todas as pessoas que por todos os meios e caminhos me fizeram chegar até aqui.

RESUMO

A ciência aberta representa um novo paradigma científico que objetiva uma cultura de abertura aos processos e transparência dos procedimentos científicos. Dentro dessa abertura, encontram-se os dados científicos de pesquisa que representam o foco da *e-Science* e que caracterizam-se tanto como insumo quanto como produto da atividade de investigação. Há diversas abordagens sobre os dados científicos de pesquisa, tais como definição do conceito de abertura de dados, licenciamento, confiabilidade, ciberinfraestrutura de repositórios de dados e também as questões comportamentais quanto ao uso, compartilhamento, reuso e visualização de dados. Há uma lacuna de estudos na área a respeito do impacto causado pelo reuso de dados por parte dos pesquisadores. Esta pesquisa busca contribuir com os estudos na área. Tem por objetivo geral identificar os fatores de impacto percebidos pelos pesquisadores quanto ao reuso de dados científicos em repositório de dados brasileiros registrados no Re3data.org. Com relação a metodologia, é uma pesquisa de abordagem mista, do tipo exploratória e descritiva. Utiliza a pesquisa bibliográfica e documental como procedimento técnico e o questionário com perguntas fechadas para levantamento dos fatores de impacto como instrumento de coleta de dados. A amostra da pesquisa é do tipo não probabilística por julgamento e formada por 336 pesquisadores com conjuntos de dados depositados nos repositórios brasileiros cadastrados no Re3data.org. Os resultados destacam o perfil dos pesquisadores e sua aderência em relação aos preceitos de compartilhamento e reuso de dados científicos de pesquisa. Os fatores de impacto sobre reuso de dados foram identificados e sintetizados em cinco categorias: condições facilitadoras, benefícios, riscos, esforços e oportunidades, categorias estas que trazem compreensões sobre as atitudes dos pesquisadores no reuso de dados. Apresenta, enquanto contribuição teórica, a identificação de fatores de impacto sobre o reuso de dados numa abordagem comportamental, e de forma prática, verifica que os repositórios brasileiros não ainda não estão integralmente adaptados para divulgação, compartilhamento e reuso de dados.

Palavras-chave: Reuso de dados científicos de pesquisa. Dados científicos de pesquisa - Impacto. Compartilhamento de dados. Repositório de dados. Ciência aberta.

ABSTRACT

Open science represents a new scientific paradigm that aims at a culture of openness to scientific processes and procedures. Within this opening, there are scientific research data that represent the focus of e-Science and which are characterized both as an input and as a product of research activity. There are several approaches to scientific research data, such as defining the concept of data openness, licensing, reliability, cyber infrastructure of data repositories and also behavioral issues regarding the use, sharing and reuse of data. There is a lack of studies in the area regarding the impact caused by the reuse of data by researchers. Thus, this research seeks to fill this gap, aiming to identify the impact factors perceived by the researchers regarding the reuse of scientific data in the Brazilian data repository registered at Re3data.org. Regarding the methodology, it's presented with a mixed approach, exploratory and descriptive. Regarding technical procedures, bibliographic and documentary research is used, using a survey with closed questions to identify impact factors. The research sample is non-probabilistic by trial and focuses on 336 researchers with datasets deposited in the Brazilian repositories registered on Re3data.org. The results highlight the profile of the researchers and their adherence to the precepts of sharing and reusing scientific research data. The impact factors on data reuse were identified and synthesized in five categories: enabling conditions, benefits, risks, efforts and opportunities, categories that bring understandings about the researchers' attitudes towards data reuse. As a theoretical contribution, it presents the identification of impact factors on data reuse in a behavioral approach, and in a practical way, it verifies that Brazilian repositories are not yet fully adapted for data dissemination, sharing and reuse.

Keywords: Data reuse. Research data - Impact. Data sharing. Data repositories. Open Science.

LISTA DE FIGURAS

FIGURA 1 - ESTRUTURA DA OPEN SCIENCE CONFORME A FOSTER	18
FIGURA 2 - DUBLIN CORE SCHEMA.....	34
FIGURA 3 - EML SCHEMA.....	35
FIGURA 4 - CICLO DE VIDA DOS DADOS.....	39
FIGURA 5 - REPOSITÓRIOS BRASILEIROS REGISTRADOS NO RE3DATA	55
FIGURA 6 - FLUXO DE REGISTRO DE REPOSITÓRIOS NO RE3DATA	57
FIGURA 7 - FLUXO DE IDENTIFICAÇÃO DE PESQUISADORES.....	65
FIGURA 8 - METADADO AUTORIA E ENDEREÇO DE E-MAIL	66
FIGURA 9 - REQUISITOS DE DISPONIBILIDADE DO RE3DATA.ORG	76
FIGURA 10 - PÁGINA DE RESUMO DE UM DATASET DA BDC/UFPR.....	79
FIGURA 11 - INDISPONIBILIDADE DO PPBIO	81
FIGURA 12 - RESUMO PPBIO DATA REPOSITORY.....	82
FIGURA 13 - EXEMPLO DE DATASET DO PPBIO DATA REPOSITORY	83
FIGURA 14 - PÁGINA RESUMO IBICT DATAVERSE NETWORK.....	86
FIGURA 15 - REDE DE CATEGORIAS DE IMPACTO.....	114

LISTA DE GRÁFICOS

GRÁFICO 1 - PERCENTUAL DE DADOS COMPARTILHADOS EM ACESSO	
ABERTO	48
GRÁFICO 2 - PRINCIPAL OCUPAÇÃO DOS RESPONDENTES.....	88
GRÁFICO 3 - PRINCIPAL LOCAL DE TRABALHO DOS PESQUISADORES.....	89
GRÁFICO 4 - SETORES DE VINCULAÇÃO PROFISSIONAL DOS	
PESQUISADORES.....	91
GRÁFICO 5 - PRINCIPAIS FINANCIADORES DE PESQUISA	92
GRÁFICO 6 - NÍVEL DE ESCOLARIDADE DOS PESQUISADORES	92
GRÁFICO 7 - TIPO DE DADO UTILIZADO PELOS PESQUISADORES	93
GRÁFICO 8 - LOCAL ONDE OS PESQUISADORES DISPONIBILIZAM SEUS	
DADOS	95
GRÁFICO 9 – PADRÕES DE METADADOS MAIS UTILIZADOS.....	102
GRÁFICO 10 - ABERTURA DOS DADOS PARA ACESSO	103
GRÁFICO 11 - EXIGÊNCIA DE PDG PELAS ENTIDADES FINANCIADORAS DE	
PESQUISA.....	103
GRÁFICO 12 – BENEFÍCIOS PERCEBIDOS SOBRE O REUSO DE DADOS.....	104
GRÁFICO 13 - OPORTUNIDADES PERCEBIDAS NO REUSO DE DADOS	107

LISTA DE QUADROS

QUADRO 1 - TIPOS DE METADADOS	40
QUADRO 2 - SEMELHANÇAS ENTRE O CICLO DE VIDA DO DADO E O CICLO DE VIDA DA GESTÃO DA INFORMAÇÃO	43
QUADRO 3 - POLÍTICA DE GESTÃO DE DADOS REQUERIDAS PELAS AGÊNCIAS DE FOMENTO BRASILEIRAS	46
QUADRO 4 - REPOSITÓRIO DE DADOS BRASILEIROS REGISTRADOS NO RE3DATA	64
QUADRO 5 - FASES DA COLETA DE DADOS	69
QUADRO 6 - OBJETIVOS ESPECÍFICOS, RESULTADOS ESPERADOS, QUESTÕES E SUPORTE TEÓRICO	70
QUADRO 7 - ANÁLISE ESTATÍSTICA DESCRITIVA: PARTE I - PERFIL DOS PESQUISADORES	73
QUADRO 8 - ANÁLISE ESTATÍSTICA DESCRITIVA: PARTE II - COMPORTAMENTO DOS PESQUISADORES QUANTO AO USO DE DADOS	73
QUADRO 9 - ANÁLISE ESTATÍSTICA DESCRITIVA: PARTE III - COMPORTAMENTO DOS PESQUISADORES QUANTO AO REUSO DE DADOS	73
QUADRO 10 - PROTOCOLO BDC/UFPR	77
QUADRO 11 - PROTOCOLO PPBIO DATA REPOSITORY	81
QUADRO 12 - PROTOCOLO IBICT DATAVERSE NETWORK	85
QUADRO 13 - DISTRIBUIÇÃO DE FREQUÊNCIA - COMO OS PESQUISADORES COLETAM E USAM DADOS	98
QUADRO 14 - DISTRIBUIÇÃO DE FREQUÊNCIAS - O ENVOLVIMENTO DAS INSTITUIÇÕES E/OU GRUPOS DE PESQUISA COM OS DADOS PRODUZIDOS PELOS PESQUISADORES	99
QUADRO 15 - CATEGORIAS E FATORES DE IMPACTO	112
QUADRO 16 - RESUMO DOS RESULTADOS OBTIDOS SOBRE AS CATEGORIAS DE IMPACTO	119

LISTA DE TABELAS

TABELA 1 - REPOSITÓRIOS EXCLUÍDOS DA ANÁLISE	68
TABELA 2 - REPOSITÓRIOS SELECIONADOS.....	68
TABELA 3 - COEFICIENTE ALFA DE CRONBACH.....	73
TABELA 4 - ÁREA DE ATUAÇÃO	90
TABELA 5 - INDISPONIBILIDADE NA PUBLICAÇÃO DE TODOS OS DADOS.....	96
TABELA 6 - COMO OS PESQUISADORES COLETAM E USAM DADOS CIENTÍFICOS DE PESQUISA (QUESTÃO 11).....	97
TABELA 7 - COMO A SUA INSTITUIÇÃO E/OU GRUPO DE PESQUISA ESTÁ ENVOLVIDA COM SEUS DADOS	98
TABELA 8 - OPINIÃO DOS PESQUISADORES QUANTO AO USO DE DADOS DENTRO DE SUAS ÁREAS DE PESQUISA	100
TABELA 9 - COMPORTAMENTO DOS PESQUISADORES QUANTO A REUSO DE DADOS	101
TABELA 10 - RISCOS PERCEBIDOS NO REUSO DE DADOS	105
TABELA 11 - ESFORÇOS PERCEBIDOS NO REUSO DE DADOS.....	105
TABELA 12 - CARACTERÍSTICAS DOS DADOS PARA REUSO	107
TABELA 13 - DESCOBERTA DE REPOSITÓRIOS DE DADOS	108
TABELA 14 - OPINIÃO SOBRE REUSO DE DADOS DENTRO DAS ÁREAS DE PESQUISA.....	109
TABELA 15 - OPINIÃO SOBRE REUSO DE DADOS	110
TABELA 17 - POLÍTICAS E/OU DOCUMENTO NORTEADORES SOBRE GESTÃO DE DADOS CIENTÍFICOS PELAS FUNDAÇÕES DE AMPARO À PESQUISA ESTADUAIS	134

LISTA DE ABREVIATURAS OU SIGLAS

API	Application Programming Interface
C3SL	Centro de Computação Científica e Software Livre
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CERIF	Common European Research Information Format
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CONFAP	Conselho Nacional Fundações Estaduais de Amparo à Pesquisa
DCC	Digital Curation Center
DIF	Directory Interchange Format
DwC	Darwin Core
EMBRAPA	Empresa Brasileira de Agropecuária
EML	Ecological Metadata Language
FAP	Fundação de Amparo à Pesquisa
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
Fiocruz	Fundação Oswaldo Cruz
HTML	HyperText Markup Language
IBICT	Instituto Brasileiro de Informação Científica e Tecnológica
ICPSR	Inter-university Consortium for Political and Social Research
INCT-CENBAM	Centro de Estudos Integrados da Biodiversidade Amazônica
INPE	Instituto Nacional de Pesquisas Espaciais
ISO	International Standards Organization
MCTI	Ministério das Ciência, Tecnologia e Inovações
METS	Metadata Encoding and Transmission Standards
NIH	National Institute of Health
NIS	National Institute of Science
NSB	National Science Board
OAI	Open Archives Initiative
OAI-ORE	Open Archives Initiative Object Reuse and Exchange
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OAIS	Open Archival Information System
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
OFK	Open Knowledge Foundation

PGD	Plano de Gestão de Dados
RDA	Research Data Alliance
RDI	Repositório Digital Institucional
RNP	Rede Nacional de Pesquisa
SciELO	Scientific Eletronic Library
SIBI	Sistema de Bibliotecas da UFPR
UFPR	Universidade Federal do Paraná
UFRGS	Universidade Federal do Rio Grande do Sul
URL	Uniform Resource Location
USP	Universidade de São Paulo
XML	Extensible Markup Language

SUMÁRIO

1 INTRODUÇÃO	18
1.1 CONTEXTUALIZAÇÃO DO PROBLEMA	19
1.2 OBJETIVOS	23
1.2.1 Objetivo geral	23
1.2.2 Objetivos específicos	23
1.3 RELEVÂNCIA E JUSTIFICATIVA PARA A REALIZAÇÃO DA PESQUISA.....	24
1.4 ESTRUTURA DA PESQUISA.....	26
2 REFERENCIAL TEÓRICO	27
2.1 CIÊNCIA ABERTA	27
2.2 DADOS CIENTÍFICOS DE PESQUISA	30
2.2.1 Características e princípios dos dados científicos	33
2.2.2 Gestão de dados científicos de pesquisa.....	38
2.3 COMPARTILHAMENTO DE DADOS CIENTÍFICOS.....	44
2.3.1 Vantagens do compartilhamento de dados.....	49
2.4 REUSO DE DADOS CIENTÍFICOS	51
2.5 REPOSITÓRIO DE DADOS DE PESQUISA	53
2.6 O IMPACTO DOS DADOS CIENTÍFICOS NAS ATIVIDADES DE PESQUISA..	59
3 METODOLOGIA	62
3.1 CARACTERIZAÇÃO DA PESQUISA.....	62
3.2 AMBIENTE E AMOSTRA DA PESQUISA	63
3.3 COLETA DE DADOS	68
3.3.1 Elaboração do instrumento de coleta de dados	69
3.3.2 Aplicação do pré-teste.....	71
3.3.3 Procedimento de coleta e análise dos dados.....	72
4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	75
4.1 DESCRIÇÃO DOS REPOSITÓRIOS DE DADOS BRASILEIROS	75
4.2 PERFIL DOS PESQUISADORES E CATEGORIZAÇÃO DOS FATORES DE IMPACTO	87
4.2.1 Perfil dos pesquisadores	88
4.2.2 Comportamento sobre dados científicos de pesquisa	93
4.2.3 Comportamento dos pesquisadores quanto ao reuso de dados.....	100
4.3 CATEGORIZAÇÃO DO IMPACTO PERCEBIDO PELOS PESQUISADORES	110

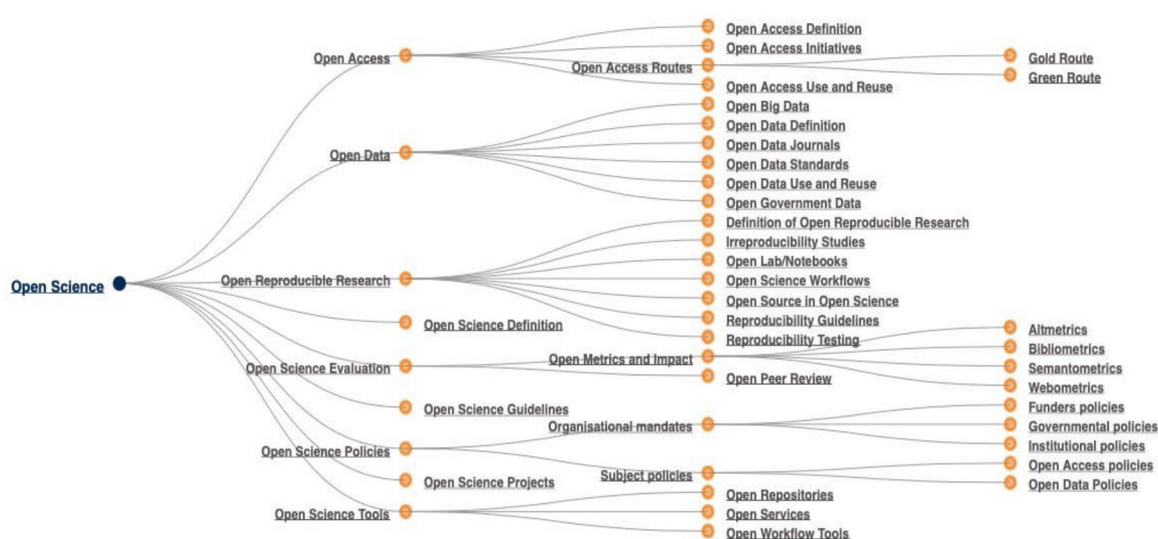
4.4 SÍNTESE DOS RESULTADOS	115
5 CONCLUSÃO	120
5.1 DIFICULDADES NA PESQUISA E SUGESTÕES DE PESQUISAS FUTURAS	
123	
REFERÊNCIAS.....	124
APÊNDICE 1 – QUESTIONÁRIO	133
APÊNDICE 2 – POLÍTICAS E/OU DOCUMENTOS NORTEADORES SOBRE	
GESTÃO DE DADOS CIENTÍFICOS PELAS FUNDAÇÕES ESTADUAIS DE	
AMPARO À PESQUISA	134
APÊNDICE 3 – CÁLCULO DO COEFICIENTE DO ALFA DE CRONBACH.....	136
APÊNDICE 4 – ESTATÍSTICA DESCRITIVA	144

1 INTRODUÇÃO

O desenvolvimento das tecnologias digitais e as ferramentas colaborativas contribuíram de forma significativa para a alteração nos modelos de comunicação da ciência. Passados vinte anos do lançamento do primeiro portal de periódicos *on-line* gratuito do Brasil, o *Scientific Eletronic Library Online* (SciELO), é possível observar a aproximação de uma nova abordagem no processo científico: a ciência aberta.

A ciência aberta é definida pela *Open Knowledge Foundation* (OKF) (s/d) mais como um princípio por meio do qual “o conhecimento científico deve ser livre para as pessoas usarem, reutilizarem e distribuírem sem restrições legais” (tradução nossa). Nesse sentido, é um conceito amplo, um “termo guarda-chuva” (ALBAGLI, 2015) ou um ecossistema que abriga em si outros princípios a exemplo do *open source* (FIGURA 1). Um desses princípios é a gestão de dados científicos de pesquisa, especificamente, o compartilhamento e reuso. Os dados científicos, nesta abordagem, são a força motriz da atividade científica, pois oferecem condições de afirmar, discordar, comprovar e refutar experimentos em qualquer área científica.

FIGURA 1 - ESTRUTURA DA OPEN SCIENCE CONFORME A FOSTER



FONTE: FOSTER Open Science (s/d).

Ainda, os dados são o principal elemento que subsidiam os resultados da pesquisa, ou seja, os dados coletados e utilizados para análise. A exigência da disponibilidade de dados está diretamente ligada à sustentabilidade da pesquisa financiada com recursos públicos (ALBAGLI, 2015). Dessa forma, as discussões

sobre a comunicação dos dados da pesquisa se intensificaram, sendo impulsionadas e viabilizadas também pelo movimento de *software* livre e acesso aberto à comunicação da produção acadêmica e científica.

A gestão de dados científicos compreende diversos aspectos relacionados aos dados, como: gestão, compartilhamento, reuso, interoperabilidade, curadoria de dados, ciberinfraestrutura, serviços de apoio, licenciamento, aspectos comportamentais, políticas públicas de informação, entre outros (PAMPEL, 2003; PIWOWAR; DAY; FRIDSMA, 2007; ARZBERGER et al., 2004; TENOPIR et al., 2011; BORGMAN, 2012; ALBAGLI, 2015; OLIVEIRA; SILVA, 2016; SAYÃO; SALES, 2016; CURTY, 2016; CURTY et al., 2017).

Os estudos brasileiros destacam vantagens e desvantagens sobre compartilhamento e reuso (ALBAGLI; MACIEL; ABDO, 2015; SAYÃO; SALES, 2016; OLIVEIRA; SILVA, 2016), todavia, não foram identificados estudos nacionais sobre o que pensam os pesquisadores brasileiros que reusam dados.

Assim, a questão do reuso de dados em si e a questão do impacto causado pelo reuso chama a atenção pela carência de estudos que interliguem esses dois temas na perspectiva nacional. A partir das pesquisas de Tenopir et al., 2011; Borgman, 2012; Albagli; Maciela; Adbo, 2015; Curty, 2016 e Curty et al., 2017 observa-se que há uma lacuna sobre estudos que identifiquem os fatores de impacto percebido pelos pesquisadores no reuso de dados.

Assim, esta pesquisa busca preencher essa lacuna, examinando as atitudes dos pesquisadores para identificar os fatores de impacto percebido por estes com relação ao reuso de dados científicos. Para tal, a pesquisa se debruça sobre o comportamento dos pesquisadores que compartilham e reusam dados científicos a partir da exploração e análise dos repositórios de dados exclusivamente nacionais registrados no diretório Re3data.org.

1.1 CONTEXTUALIZAÇÃO DO PROBLEMA

As agências de fomento em pesquisa reconhecem a importância da publicação dos dados e exigem uma adequada gestão destes dados durante o processo de investigação, evidenciando a necessidade de assegurar meios e mecanismos para preservação e reutilização dos mesmos. O *National Institute of*

Health (NIH) criou sua política de acesso em 2005 e em 2007 foi aprovada uma legislação nacional que exigia da entidade o acesso aberto de todas as pesquisas financiadas pelo instituto, sendo que todos os artigos deveriam ser depositados no repositório institucional *PubMedCentral* (GRIGG, 2015).

A *National Science Foundation* (NSF) requer, desde 2011, que as pesquisas financiadas sejam acompanhadas de um Plano de Gestão de Dados (PGD) e a disponibilidade desses dados ocorra em acesso aberto (GRIGG, 2015). O programa *Europeu Horizon 2020* (H2020), desde 2013, também faz essa exigência.

O Reino Unido optou pela criação em 2004 de um consórcio integrado por diversas universidades para centralizar a criação de políticas institucionais, capacitações e desenvolvimento de habilidades com dados - esse consórcio foi denominado *Digital Curation Center* (DCC) e é um centro de referência para a gestão de dados científicos de pesquisa. O DCC merece destaque pela colaboração com outras universidades norte-americanas na elaboração do DMPTool¹ uma ferramenta para criação e gerenciamento de PGD, e também por seu destaque no estudo de padrões para descrição de dados (metadados).²

Na perspectiva brasileira, a Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) requer dos pesquisadores a elaboração de um PGD como item obrigatório para atendimento das solicitações de recursos para o financiamento de projetos.

Como principal fomentador da pesquisa e da formação de pesquisadores no Brasil, o Conselho Nacional de Desenvolvimento Cultural e Tecnológico (CNPq) não possui uma política de gestão de dados. Contudo, algumas das chamadas públicas lançadas em 2018 e 2019³ possuem recomendação para que as publicações originadas no âmbito das pesquisas financiadas por este Conselho sejam publicadas “preferencialmente” em revistas de acesso aberto. Interessante observar que as chamadas públicas do CNPq orientam a obrigatoriedade de agradecimento ao Conselho para as publicações financiadas como forma de tornar clara a fonte do recurso de financiamento. O que parece justo à medida que o financiamento é

¹ Disponível em: <https://dmptool.org>. Acesso: 10 mar. 2019.

² Disponível em: <http://www.dcc.ac.uk/resources/metadata-standards>. Acesso em: 10 mar. 2019.

³ Chamadas públicas n. [38/2018](#), [39/2018](#), [40/2018](#) e [01/2019](#). Acesso em: 10 mar. 2019.

público e confere ares de transparência na aplicação de investimentos do órgão. Por outro lado, a mesma regra de publicidade e transparência não é aplicada de forma mandatória às publicações e aos dados, já que a recomendação é flexível quanto à disponibilização em acesso aberto de publicações financiadas pelo Conselho, o que de certa forma, não confere caráter de abertura e transparência para o órgão quanto às publicações oriundas de projetos subsidiados.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), fundação que atua na organização e consolidação da pós-graduação *strictu sensu* no país ainda não definiu orientações sobre política de gestão dados. As iniciativas em acesso aberto são discutidas internamente pelas universidades e respectivos programas de pós-graduação.

A Fundação Oswaldo Cruz (Fiocruz) possui uma política de acesso aberto criada em 2014 para o âmbito das publicações e não para os dados. Em 2017, a entidade publicou o “Livro Verde - ciência aberta e dados abertos” que apresenta uma pesquisa sobre o contexto da ciência aberta na Europa com vistas a subsidiar diretrizes de acesso aberto para a Fiocruz (SANTOS, et al., 2017).

No entanto, ainda que haja crescente exigência pela abertura e compartilhamento de dados e fomento ao reuso, os pesquisadores enfrentam diversas dificuldades e desafios diante desta prática. Uma das dificuldades destacadas por Jorge e Albagli (2018, p. 417) e Arzberger et al. (2004) diz respeito sobre a diferenciação entre abertura e compartilhamento de dados, observando que a diferença entre um e outro reside na condição de acesso e uso, reutilização, reprodução e transparência dos dados, já que “dados compartilhados nem sempre seguem os princípios dos dados abertos”. O conceito de abertura de dados no contexto deste projeto é definido de acordo com os Princípios FAIR (que de maneira geral caracterizam a abertura de dados por meio de 4 princípios: *findable* (localizáveis), *accessible* (acessíveis), *interoperable* (interoperável) e *reusable* (reutilizável) (WILKINSON, 2016; AVENTURIER, 2017; HENING, et al., 2019).

Já as dificuldades sobre o compartilhamento residem em aspectos epistemológicos sobre a tipologia de dados gerados nos processos de investigação (BORGMAN, 2012; SAYÃO; SALES, 2014), sobre a falta de definição do que é um dado científico de pesquisa (TENOPIR et al., 2011; FRIEDERICK, 2016; OLIVEIRA; SILVA, 2016), falta de confiabilidade nos dados (CURTY et al., 2017), e questões de infraestrutura dos repositórios de dados (PAMPEL et al., 2013; VIERKANT et al.,

2014; SAYÃO; SALES, 2016; CATLIN et al., 2018; WU et al., 2019) e também questões comportamentais dos pesquisadores sobre compartilhamento e reuso de dados (TENOPIR et al., 2011; BORGMAN, 2012; CURTY, 2015; CURTY et al., 2017).

Então, diante das exigências das agências financiadoras e considerando os desafios e dificuldades da abertura, compartilhamento e reuso de dados, depreende-se a relevância de investimentos para promover a ciência aberta, não apenas para dar visibilidade aos resultados das pesquisas, mas também na disponibilidade dos dados que fundamentam estes resultados em artigos e textos científicos financiados por instituições de ensino e agências de fomento.

O reuso de dados é considerado uma tendência para acelerar o avanço da ciência de forma mais equitativa. Beneficia a sociedade com o uso dos recursos e investimentos de forma inteligente e economicamente viável, preservando os dados como ativos úteis para novas gerações.

O compartilhamento e reuso de dados científicos de pesquisa é uma temática em ascensão dentro do amplo escopo da ciência aberta e carece ainda de estudos complementares sobre os demais aspectos que interferem e/ou influenciam o reuso dos dados científicos (ALBAGLI; MACIEL; ABDO, 2015).

Um desses aspectos que merece destaque é o fator comportamental em reuso de dados. Curty (2016) apresentou uma pesquisa a qual identificou 25 fatores individuais sociais e tecnológicos que influenciam a intenção e o comportamento do reuso de dados entre os cientistas sociais. Borgman (2012), Curty (2016) e Curty et al., 2017) destacam que no desafio comportamental em relação à reusabilidade de dados há uma lacuna sobre estudos que identifiquem os fatores de impacto quanto ao reuso de dados científicos, bem como as motivações para compartilhamento e reuso desses dados.

Nesse sentido, esta investigação busca responder a seguinte questão: **quais são os fatores de impacto percebidos pelos pesquisadores quanto ao reuso de dados científicos?**

A identificação apurada dos desafios e dificuldades observados na literatura da área sobre o reuso de dados pode trazer compreensão sobre quais as atitudes e os fatores de impacto que os pesquisadores observam ao reutilizarem dados científicos em suas práticas de pesquisa. A identificação dos fatores de impacto pode estar relacionada diretamente com os benefícios, vantagens e/ou

oportunidades subjacentes que podem advir da publicação de dados científicos em repositórios, seja na forma redução de recursos, tempo, custos, citação, visibilidade, influência social, entre outros.

1.2 OBJETIVOS

Os objetivos são divididos em objetivo geral e quatro objetivos específicos.

1.2.1 Objetivo geral

Identificar os fatores de impacto percebido pelos pesquisadores sobre o reuso de dados científicos nos repositórios brasileiros registrados no Re3data.

1.2.2 Objetivos específicos

Os objetivos específicos que apoiarão o desenvolvimento da pesquisa são:

- a) identificar o impacto sobre reuso de dados científicos na literatura;
- b) identificar e descrever os repositórios de dados brasileiros registrados no Re3data.org;
- c) identificar os pesquisadores que apresentam *datasets* depositados nos repositórios brasileiros; e
- d) categorizar os fatores de impacto percebido pelos pesquisadores no reuso de dados científicos.

Ao final da pesquisa, busca-se apresentar uma contribuição teórica e prática com relação ao reuso de dados científicos de pesquisa. A contribuição teórica no sentido de contribuir para estudos sobre atitudes e comportamento dos pesquisadores com relação a reuso de dados na perspectiva nacional, contribuindo com os estudos na área de ciência aberta especificamente em dados científicos de pesquisa. A contribuição prática aparece no sentido de indicações aos gestores de repositórios com relação ao cumprimento das recomendações dos princípios de abertura de dados como forma a tornar os repositórios adaptados à disponibilização e compartilhamento de dados para promoção do reuso.

1.3 RELEVÂNCIA E JUSTIFICATIVA PARA A REALIZAÇÃO DA PESQUISA

No contexto da ciência aberta, a gestão eficaz de dados científicos se apresenta como um importante elemento propulsor de novos estudos, além de favorecer a transparência e a divulgação da ciência para a sociedade, o que justificativa cientificamente o desenvolvimento desta pesquisa.

Como justificativa pessoal, destaca-se o envolvimento da pesquisadora com as iniciativas de acesso aberto desde a graduação e também a gestão dos dados científicos de pesquisa em sua prática profissional como bibliotecária nesta universidade.

O interesse pelo tema de pesquisa surgiu por meio de uma dúvida levantada no desenvolvimento da atividade profissional desta autora sobre qual seria o impacto percebido pelos pesquisadores ao reutilizarem dados científicos, em outras palavras, que tipo de impacto em sua atividade acadêmica/científica é percebida pelos pesquisadores ao reutilizarem dados científicos. Eles esperam impactos positivos, negativos, a ampliação de sua visibilidade, recompensas ou apenas altruísmo científico? Então, em levantamento bibliográfico inicial verificou-se um campo de investigação com uma lacuna passível de maior detalhamento sobre reuso de dados científicos. Dessa forma, considerou-se relevante buscar familiaridade e apropriação de conhecimentos no assunto para possibilitar informação científica no campo da Gestão da Informação sobre a lacuna identificada e que possa contribuir com estudos relacionados ao reuso de dados de pesquisa e assim beneficiar a gestão de repositórios de dados em diferentes domínios do saber.

Ainda que não seja a forma exclusiva, o acesso aberto tem sido uma prática cada vez mais estimulada pelas agências de financiamento e pela própria comunidade científica. São iniciativas da própria comunidade, como o movimento pelo *software* livre de código aberto, posteriormente o movimento pelo acesso aberto às publicações científicas que, em conjunto com as ferramentas e protocolos técnicos desenvolvidos por essas mesmas comunidades têm sustentado a execução de políticas de acesso aberto e compartilhamento de dados científicos de pesquisa, o que justifica de forma social a existência desta pesquisa. Ainda como justificativa social, o reuso de dados científicos de pesquisa possibilita economia de recursos valiosos no ciclo da pesquisa.

Como justificativa acadêmica, o problema de pesquisa apresentado nesta visa contribuir para uma compreensão mais apurada sobre o reuso de dados científicos na visão do pesquisador-cientista, pretendendo somar-se aos demais estudos sobre dados científicos e possibilidades de compartilhamento e reuso e as práticas da *e-Science* (OLIVEIRA; SILVA, 2016).

Como justificativa institucional, esta pesquisa alinha-se aos objetivos do Programa de Pós-graduação em Gestão da Informação (PPGGI/UFPR) no sentido de explorar as relações epistemológicas e de problemática na área da Ciência da Informação e Tecnologia, a partir da discussão em torno da gestão de dados científicos, unidade nuclear dentro dos processos de Gestão da Informação e parte também dos processos de comunicação científica.

A partir da observação de que os dados têm tido uma relevância ainda maior na última década na geração do conhecimento científico, sua produção exponencial, alcançada pela disponibilidade de tecnologias de informação e comunicação, pela internet, e por recursos de geração de grande volume de dados tem lhe conferido o poder de mudar a dinâmica da própria rede (BUFREM et al., 2016; CURTY, 2016).

Os dados de pesquisa em si não são um elemento novo no processo de produção do conhecimento (TENOPIR, et al., 2011). Os dados de pesquisa que eram produzidos e esquecidos, desorganizados e não acessíveis nos arquivos pessoais dos pesquisadores ou das instituições e, na medida em que são tratados, padronizados, preservados e disponibilizados de forma gratuita, constituem um novo meio para se fazer ciência, o que também pode ser compreendido como justificativa social e de certa forma econômica.

O potencial cognitivo dos dados redesenha, através do reuso, os fluxos tradicionais de comunicação científica, estabelecendo novos padrões de socialização e de trabalho cooperativo independente de barreiras geográficas e disciplinares. O valor do dado de pesquisa está diretamente relacionado à possibilidade de uso e ao seu potencial de ser reinterpretado em outras áreas e contextos diferentes da que originalmente o gerou (SAYÃO; SALES, 2014, p. 80).

Sob esse aspecto, a justificativa teórica e de relevância quanto ao problema de pesquisa é que a esta investigação busca aproximar-se de uma compreensão sobre o *ethos* científico nos princípios “comunalismo” e “universalismo” das Normas de Merton (1973) para a prática da boa pesquisa científica (KIM; STANTON, 2015, p. 13). Compreender os fatores que impactam no reuso de dados no contexto

nacional pode promover compreensão sobre suas dificuldades, expectativas e necessidades que podem servir à elaboração de políticas públicas sobre informação científica a nível institucional e nacional.

Outra justificativa teórica para a abordagem do problema está na compreensão do contexto institucional ou de área do conhecimento que tem influência direta sobre os comportamentos de pesquisadores no compartilhamento e reuso de dados além dos aspectos individuais e tecnológicos inerentes às práticas de gestão de dados.

1.4 ESTRUTURA DA PESQUISA

A pesquisa está dividida em cinco seções. A primeira seção introduziu a pesquisa, apresentando o tema, a problemática de pesquisa, a justificativa, os objetivos geral e específicos. A segunda seção apresentou o referencial teórico. A terceira seção explicou os métodos utilizados na pesquisa: caracterização, abordagem, ambiente, amostra, coleta e análise de dados. Na quarta seção foram apresentados os resultados obtidos. Na quinta seção foram discutidos os resultados e apresentada uma categorização dos fatores de impacto identificadas junto aos pesquisadores por meio do instrumento de coleta de dados.

2 REFERENCIAL TEÓRICO

O referencial teórico teve embasamento no levantamento bibliográfico trabalhado na pesquisa, de forma a dar suporte no alcance do objetivo geral e objetivos específicos, sendo: ciência aberta, dados científicos de pesquisa, compartilhamento e reuso de dados e repositório de dados.

2.1 CIÊNCIA ABERTA

Não é novo que a produção do conhecimento científico tem como premissa o uso dos conhecimentos produzidos anteriormente para corroborar ou refutar outro resultado. Essa relação de debate científico está explicada detalhadamente em Latour (2010). A Ciência da Informação e a Biblioteconomia também expõem essas relações por meio dos estudos bibliométricos. Entretanto, o uso dos dados brutos de uma pesquisa como insumo que geram novas inferências em outro trabalho é um fenômeno mais recente.

De forma a viabilizar o acesso aberto à produção científica, duas mudanças contribuíram com a democratização do acesso aos textos produzidos nas pesquisas. A primeira diz respeito ao movimento pelo *software* livre de código fonte aberto que viabilizou a infraestrutura tecnológica necessária ao livre acesso. Iniciado ainda na década de 1970, após diversos desdobramentos do movimento do *software* livre, foi o que favoreceu o desenvolvimento das ferramentas para publicação e disponibilização da produção científica de forma aberta e gratuita (EUROPEAN COMMISSION, 2006). A segunda, mais recentemente, refere-se às mudanças políticas das agências de fomento que tem promovido um ambiente social favorável à democratização da ciência, na medida em que reduz ou anula os custos de acesso à pesquisa financiada com recursos públicos e/ou privados.

Segundo a *Open Science and Research Initiative*, a principal premissa da pesquisa aberta é a democratização da ciência (OPEN SCIENCE AND RESEARCH INITIATIVE, 2014, p. 2). A perspectiva da ciência aberta transcende então o acesso aos resultados de pesquisas. Além do acesso aberto às publicações, os dados de pesquisa abertos, o *software* de fonte aberta, a colaboração aberta, a revisão por pares aberta, as anotações de pesquisa abertas, os recursos e práticas educacionais abertos, as monografias, as dissertações e teses abertas, a ciência

cidadã e o financiamento coletivo de pesquisa também estão nas fronteiras da ciência aberta (BUENO DE LA FUENTE, 2016).

Nessa acepção, a prática da ciência é realizada de tal forma que outros possam colaborar e contribuir e sempre que os dados da pesquisa, notas de laboratório e outros processos de pesquisa estão disponíveis gratuitamente, em termos que permitam a reutilização, a redistribuição e a reprodução da pesquisa com seus dados e também dos métodos subjacentes.

A ciência aberta faz uso intensivo das tecnologias que possibilitam a colaboração *on-line* ampla para atividades científicas em todos os campos do conhecimento, de forma aberta e transparente, permitindo maior envolvimento social e favorecendo a cidadania. E diante dessa concepção, a ciência aberta pode ser considerada um importante mecanismo para a viabilidade da pesquisa científica contemporânea que utiliza intensivamente dados digitais de pesquisa (SAYÃO; SALES, 2016). Com isso, a gestão de dados se torna relevante no contexto das instituições de pesquisa científica e dos financiadores da ciência.

Para Albagli (2015, p. 14), “a ciência aberta promove o aumento dos estoques de conhecimento público, propiciando não apenas a ampliação dos índices gerais de produtividade científica e de inovação, como também taxas de retornos sociais dos investimentos em ciência e tecnologia”.

Como benefícios científicos e socioeconômicos da ciência aberta destacam-se (OECD, 2007, p. 10; TENOPIR et al., 2011; SAYÃO; SALES, 2014, p. 80):

- Incentiva a diversidade de análise e opinião;
- Promove novos tipos de pesquisa;
- Possibilita a aplicação de ferramentas automatizadas on-line de descoberta de conhecimento;
- Permite a verificação de resultados prévios;
- Torna possível o teste de hipóteses e de métodos de coleta de dados e de mensuração;
- Possibilita a exploração, por outros pesquisadores, de tópicos não previstos pelos pesquisadores iniciais;
- Permite a criação de novos conjuntos de dados, de informações e de conhecimentos quando os dados de múltiplas fontes são combinados; e
- Promove a pesquisa interdisciplinar, intersetorial, interinstitucional e internacional.

Os benefícios acima destacados não são exclusivos à ciência aberta, pois processos de colaboração, validação e reanálises são intrínsecos à atividade de pesquisa. Entretanto, as práticas de abertura desse novo fazer científico tem foco nos dados de pesquisa e buscam explorá-lo de forma aberta, irrestrita e gratuita com forte suporte na tecnologia no sentido de promover sustentabilidade à atividade científica e comunicação sem barreiras de seus resultados. “É uma ciência voltada para grandes conjuntos de dados responsáveis por um amplo conjunto de simulações, metodologias e modelos efetivados em diversas áreas do conhecimento” (OLIVEIRA; SILVA, 2016, p. 11).

A partir dos anos 2000, as políticas de comunicação de dados científicos impulsionaram a disponibilidade e compartilhamento dos dados obtidos com recursos públicos enquanto fonte de financiamento científico. O *National Institute of Health* (NIH) nos Estados Unidos e a OCDE criaram requisitos de publicização e acesso aos dados resultantes das pesquisas que financiadas com recursos públicos de modo a iniciar um processo transparência aos investimentos (OCDE, 2007).

No Brasil, destacam-se algumas iniciativas envolvendo a ciência aberta nas áreas de agropecuária, biodiversidade e saúde (SAYÃO; SALES, 2013). Entidades como a Empresa Brasileira de Agropecuária (EMBRAPA), Instituto Nacional de Pesquisas Espaciais (INPE) e a Fundação Oswaldo Cruz (Fiocruz) possuem repositórios que armazenam e disponibilizam de forma aberta suas publicações, relatórios e dados sobre as atividades desses órgãos. A Fiocruz destaca-se como pioneira na abertura de dados em situações de emergência em saúde pública como no caso do vírus Zika ao disponibilizar os dados científicos das pesquisas e permitir o compartilhamento, reuso e colaboração das pesquisas (JORGE, 2018).

Embora existam iniciativas consistentes, o Brasil ainda não formalizou uma política pública de informação orientada à ciência aberta. Oliveira e Silva (2016) apresentam estudo no qual elencam cinco dimensões que poderiam substanciar uma política de informação focada na *e-Science*: dimensão epistemológica, dimensão política, dimensão ético-legal-cultural, dimensão morfológica e dimensão tecnológica. As duas primeiras dimensões focam na conceituação sobre o que seriam dados científicos e as estruturas organizacionais e políticas de enquadramento; a dimensão ético-legal-cultural é focada em questões de apropriação do conhecimento no sentido de garantir os dados de pesquisa como “bem intelectual público e comum a todos” passando ainda por questionamentos a

cerca de autoria dos dados; a dimensão morfológica concentra-se na dinâmica e modelos metodológicos e tecnológicos inter-relacionados no gerenciamento dos dados científicos. Essa dimensão considera modelos de fluxos de dados (ciclo de vida dos dados) interligados e retroalimentados; e por fim, a dimensão tecnológica apresenta os pressupostos tecnológicos de uma ciberinfraestrutura complexa para a execução do ciclo de vida dos dados.

A pensar no dado científico como unidade nuclear do processo de investigação, a próxima seção buscará dimensionar o dado científico de pesquisa, suas características e manifestações.

2.2 DADOS CIENTÍFICOS DE PESQUISA

Cada revolução científica carrega em si uma alteração no método de trabalho e no comportamento dos pesquisadores (KUHN, 1962). O desenvolvimento e ampliação das redes digitais de comunicação, o surgimento e o investimento constante em ciberinfraestrutura aliado a uma cultura emergente de abertura da prática científica e da divulgação aberta de seus resultados são características marcantes da última revolução científica, denominada Quarto Paradigma da Ciência ou *e-Science* (HEY; TANSLEY; TOLLE, 2009; CORDEIRO et al., 2013; BUFREM et al., 2016; OLIVEIRA; SILVA, 2016).

Conforme Gray (2007 citado por HEY; TANSLEY; TOLLE, 2009), a transição gradual do fazer científico, antes aportado no caráter de componente teórico (por meio de modelos e generalizações) para a tendência computacional nas últimas décadas, determinou de sobremodo a dependência dos aparatos tecnológicos no campo científico. Essa transição apresenta como elementos o desenvolvimento de novas disciplinas como a *e-Science*, “*Data Science*” (SMITH, 2006, p. 163) ou “*Data-intensive Science*” (LYNCH, 2009, p. 177).

A *e-Science* tem como característica o suporte da ciberinfraestrutura (redes e equipamentos) e tem como foco central a obtenção e a exploração do dado científico (HEY; TANSLEY; TOLLE, 2009). Dessa forma, na mudança de paradigma do fazer científico, observa-se um novo enfoque sobre o dado.

Para Cordeiro et al. (2013, p. 73):

O acelerador de partículas *Large Hadron Collider* (LHC), da Organização Europeia para Pesquisa Nuclear (Cern), captura 25 petabytes (quatrilhões) de dados todos os anos [...]. É impossível processar essa enorme quantidade de dados “manualmente”, o processamento precisa ser obrigatoriamente feito por um software. A informação precisa ser armazenada em grandes bancos de dados. A análise desses dados tem de utilizar um ferramental estatístico avançado, codificado na forma de programas de computador que consomem, filtram, manipulam, transformam e consolidam esses dados como o objetivo de extrair alguma informação relevante.

Nesse contexto, os dados científicos constituem o elemento central que subsidia os resultados da pesquisa, pois são os dados coletados e utilizados para análise que serão posteriormente tratados e condensados em forma de resultados divulgados nos textos científicos.

O destaque sobre a exploração do dado científico manifesta-se nas inúmeras possibilidades de geração de novos conhecimentos, novas abordagens e a possibilidade de alargamento das redes de colaboração e compartilhamento (OLIVEIRA; SILVA, 2016).

Ademais sobre a importância dos dados, ainda não há consenso definitivo na literatura sobre o conceito de dado. Para Frederick (2016, p. 11), dados, de forma mais elementar são:

a força vital da ciência e tecnologia [...] os dados são essencialmente o que faz com que a tecnologia funcione na vida real [...]. Padrões e irregularidades nos dados podem ser a chave para desbloquear o conhecimento sobre causas e curas para doenças. Os dados também são usados para medir o status de nossa saúde e calcular quanto dinheiro devemos ao governo todos os anos. Apesar disso, “dado” é relativamente um conceito obscuro que tem uma ampla definição e aplicação” (tradução nossa).

Dessa forma, o dado é visto como uma representação de eventos que ocorreram ou ocorrem no mundo real, como por exemplo as alterações na pressão sanguínea e a quantidade de aves observadas em determinado período e localização geográfica. No entanto, para esse dado fazer sentido, ele deve estar dentro de um contexto e capaz de ser processado por seres humanos e máquinas (FREDERICK, 2016). Na visão da autora, o progresso da humanidade é possível devido às práticas científicas construídas em torno dos processos de coleta e análise de dados.

Na mesma linha, Borgman (2012, p. 1061) observa que “os dados podem existir somente no olho de quem vê” e isso pressupõe que o reconhecimento de uma observação, artefato ou registro constitui o que é um dado. A autora destaca também a dificuldade em definir precisamente o que é dado devido às diversificadas formas físicas e digitais em que pode se manifestar nas diversas áreas do conhecimento.

Já a definição da OECD é mais precisa e descreve o dado científico como “registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fonte primária para a pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico” (OECD, 2007, p.13, tradução nossa).

Na mesma linha, Silva (2019, p. 22) afirma que

os dados são a matéria-prima gerada a partir da concepção/coleta/observação dos fenômenos da realidade (números, caracteres, símbolos, imagens, sons, ondas eletromagnéticas, bits, etc.). Constituem os componentes básicos a partir dos quais se concebe a informação e o conhecimento. São geralmente de natureza representativa (p. ex. de medições de variáveis, tais como a idade, a altura, o peso, a cor, a pressão arterial, a opinião, os hábitos, a localização de uma pessoa, etc.).

As quatro tentativas de definição acima assumem que uma característica primordial do dado é fornecida pela observação e registro proveniente de questionamentos a cerca da realidade que leva em consideração a vida útil e origem do dado. E para distinguir as diferentes categorias que um dado científico pode assumir, conforme as distintas áreas do conhecimento em que se manifesta, foi criada em 2005 uma tipologia de dados pela *National Science Board* (NSB) que é um conselho da fundação nacional de ciências (*National Science Foundation*) dos Estados Unidos.

As diferentes tipologias de dado categorizados pela NSB considera como características dos dados: a natureza, a reprodutibilidade, e nível de processamento em que estes podem ser submetidos. Os tipos são: observacionais, computacionais e experimentais (NSB, 2005; BORGMAN, 2012; SAYÃO; SALES, 2014; SILVA, 2019).

Os dados observacionais são obtidos por meio de observações diretas, de natureza instantânea e por isso são qualificados como registros históricos que não podem ser coletados uma segunda vez. Os dados computacionais são provenientes

da execução de modelos computacionais e de simulação, e embora a recomendação seja preservar os dados de saída, há também a necessidade de preservar o ambiente desse dado (*hardware* e *software*), de maneira a não inviabilizar o acesso e manipulação posteriores. Os dados experimentais são nascidos de situações controladas de experimentos e que podem ser precisamente reproduzidos por *scripts*, porém, tal como os dados de simulação, requer a criação e/ou emulação do ambiente e insumos para reprodução do dado (SAYÃO; SALES, 2014).

Assim, é possível dizer que um dado científico é um dado primário que pode ser tratado tanto como insumo quanto subproduto dos processos de pesquisa. E dessa maneira, é importante que este esteja em acesso aberto de forma a permitir validação, reuso, novas configurações e análises, compartilhamento e colaboração, retroalimentando não só a pesquisa primária em si quanto as derivadas.

Ademais das definições generalistas sobre dado científico, é necessário destacar quais são as características, critérios e princípios que garantem a qualidade dos dados para publicação, compartilhamento e reuso, o que será apresentado na seção seguinte.

2.2.1 Características e princípios dos dados científicos

Dados científicos abertos “são dados publicados em um formato legível por máquina e sem restrição de licenças, patentes ou mecanismos de controle, de modo a estarem livremente disponíveis para serem utilizados e redistribuídos à vontade” (SANTOS, 2017, p. 122). No entanto, para que eles estejam realmente abertos, devem manter uma identificação consistente no decorrer do tempo que permita interpretá-los de forma coerente para novas pesquisas. Os metadados detalhados de significação e contexto são fundamentais, considerando a complexidade e a heterogeneidade dos dados científicos em diferentes domínios, além de estruturas que permitam interoperabilidade entre sistemas e as devidas atribuições de licenças.

Metadados se caracterizam por serem “(...) dado que descreve atributos de um recurso, caracteriza suas relações, apoia sua descoberta e uso efetivo e existe em um ambiente eletrônico. Usualmente, consiste em um conjunto de elementos, cada qual descrevendo atributo do recurso, seu gerenciamento ou uso” (VELLUCI, 1998, p. 192 citado por CAMPOS, 2007, p. 19).

O *Dublin Core* (DC) é o padrão de metadados amplamente utilizado pela maioria repositórios digitais sendo um recurso semântico para descrição de recursos eletrônicos com objetivo de promover a recuperação de dados e objetos na *web*. É um padrão simples, flexível e interoperável que pode agregar elementos qualificadores para enriquecer a descrição dos dados (DCMI, s/d, não paginado). O padrão está presente nos repositórios BDC/UFPR e IBICT *Dataverse Network*, repositórios que formam o ambiente desta pesquisa.

FIGURA 2 - DUBLIN CORE SCHEMA

```
<?xml version="1.0" encoding="UTF-8" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns="http://purl.org/dc/elements/1.1/" targetNamespace="http://purl.org/dc/elements/1.1/" attributeFormDefault="unqualified">
  <xs:annotation>
    <xs:documentation xml:lang="en">...</xs:documentation>
  </xs:annotation>
  <xs:import namespace="http://www.w3.org/XML/1998/namespace" schemaLocation="http://www.w3.org/2001/03/xml.xsd"></xs:import>
  <xs:element name="title" type="elementType"/>
  <xs:element name="creator" type="elementType"/>
  <xs:element name="subject" type="elementType"/>
  <xs:element name="description" type="elementType"/>
  <xs:element name="publisher" type="elementType"/>
  <xs:element name="contributor" type="elementType"/>
  <xs:element name="date" type="elementType"/>
  <xs:element name="type" type="elementType"/>
  <xs:element name="format" type="elementType"/>
  <xs:element name="identifier" type="elementType"/>
  <xs:element name="source" type="elementType"/>
  <xs:element name="language" type="elementType"/>
  <xs:element name="relation" type="elementType"/>
  <xs:element name="coverage" type="elementType"/>
  <xs:element name="rights" type="elementType"/>
  <xs:group name="elementsGroup">
    <xs:sequence>
      <xs:choice minOccurs="0" maxOccurs="unbounded">
        <xs:element ref="title"/>
        <xs:element ref="creator"/>
        <xs:element ref="subject"/>
        <xs:element ref="description"/>
        <xs:element ref="publisher"/>
        <xs:element ref="contributor"/>
        <xs:element ref="date"/>
        <xs:element ref="type"/>
        <xs:element ref="format"/>
        <xs:element ref="identifier"/>
        <xs:element ref="source"/>
        <xs:element ref="language"/>
        <xs:element ref="relation"/>
        <xs:element ref="coverage"/>
        <xs:element ref="rights"/>
      </xs:choice>
    </xs:sequence>
  </xs:group>
  <xs:complexType name="elementType">
    <xs:simpleContent>
      <xs:extension base="xs:string">
        <xs:attribute ref="xml:lang" use="optional"/>
      </xs:extension>
    </xs:simpleContent>
  </xs:complexType>
</xs:schema>
```

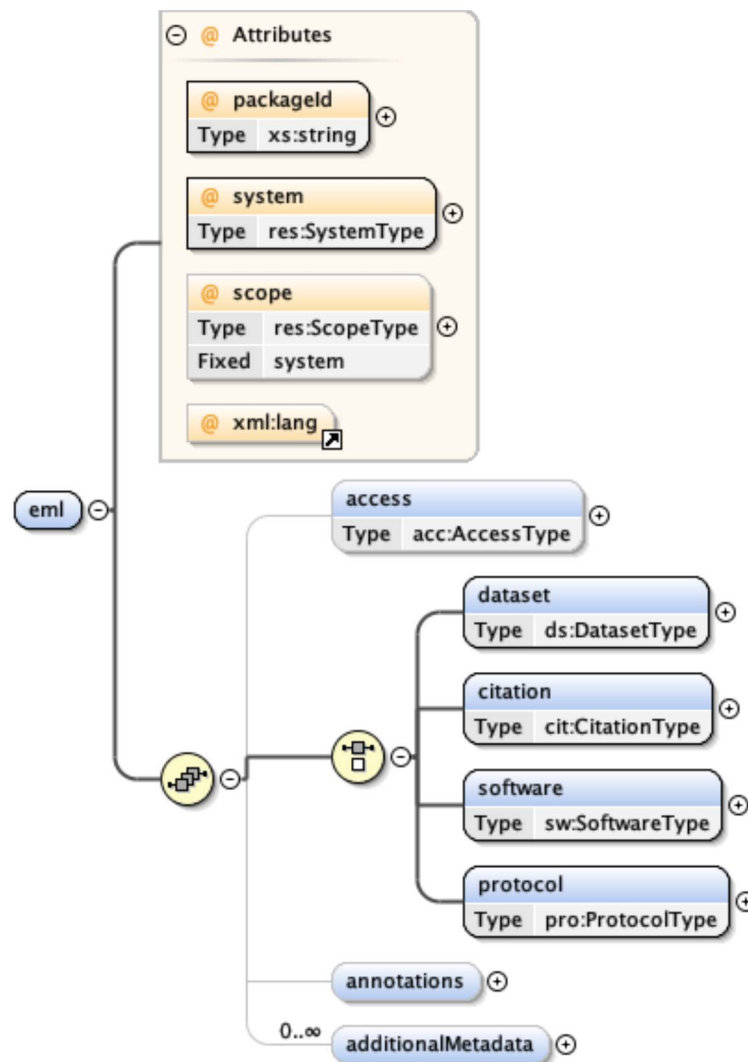
FONTE: DublinCore.org (online)⁴

O *Ecological Metadata Language* (EML)⁵ é um padrão de metadados desenvolvido especificamente para gestão de dados nas áreas de Ciências

⁴ DC schema. Disponível em: <https://dublincore.org/schemas/xmls/simpledc20021212.xsd>. Acesso em 27 fev. 2020.

Ambientais, Biológicas e da Terra, embora seja possível o uso em diversas áreas do conhecimento. Com o EML é possível a descrição e identificação de pacotes de dados, extensão espacial, temporal, taxonômica e temática dos dados, além de *tags* específicas para descrição de métodos e protocolos utilizados na construção do dado (JONES et al., 2019). O EML possui módulos para descrição desde o projeto de pesquisa até a unidade nuclear do dado científico, conforme figura abaixo:

FIGURA 3 - EML SCHEMA



FONTE: Ecological Metadata Language.org (*on-line*).

Em relação à abertura de dados, salienta-se a caracterização adotada pela OKF, entidade que atua em diversas frentes da ciência aberta, que estabeleceu em 2009 condições que definem a abertura do acesso e uso dos dados, sendo elas:

⁵ EML schema. Disponível em: <https://eml.ecoinformatics.org/schema/index.html>. Acesso em: 27 fev. 2020.

1. É livre para uso e reuso;
2. É distribuído sem restrições;
3. Utiliza licença que determina que dados derivados compartilhem a mesma licença de abertura;
4. É disponível em todo o seu conjunto;
5. Quando possuir custo de reprodução, este deve ser razoável;
6. Preferencialmente apto para acesso por meio da internet e sem custos; e
7. É disponível em formato conveniente e modificável.

Mas, além de características de abertura, para que a gestão de dados seja eficaz e os dados tratados se mantenham íntegros para o reuso com segurança, existem recomendações de boas práticas em gestão, dentre elas, destacam-se os princípios FAIR.

Os princípios FAIR, do acrônimo *Findable* (encontrável), *Accessible* (acessível), *Interoperable* (interoperável) e *Reusable* (reutilizável) representam “um conjunto mínimo de princípios orientadores e práticas aceitas pela comunidade para que os produtores e os usuários, humanos ou computadores, [possam] usar mais facilmente os dados e citar eles corretamente.” (AVENTURIER, 2017, não paginável, tradução nossa). Surgido, inicialmente, como recomendação de boas práticas para a ciência aberta, observam-se aplicações desses princípios na comunidade europeia para que sejam estabelecidos como requisitos à gestão de dados.

Ressalta-se que estes princípios não são ‘fechados’, segundo os próprios autores mencionam, eles devem ser “inteligentemente abertos” para que possam respeitar as particularidades das diferentes áreas do saber científico. Eles orientam para que as escolhas na gestão de dados possam ser assertivas e permitam o compartilhamento e reuso no contexto da ciência aberta, respeitando as questões éticas e legais.

“Os princípios do FAIR fornecem diretrizes, não requerem tecnologias específicas e permitem que as comunidades de partes interessadas definam padrões FAIR específicos e desenvolvam métricas para quantificá-los”. (DUMONTIER; WESLEY, 2018, p. 39, tradução nossa).

Constituem-se em princípios FAIR – aqueles que “têm por objetivo tornar os dados de pesquisa (por intermédio de metadados) encontráveis, acessíveis, interoperáveis e reutilizáveis (*findable, accessible, interoperable and reusable*), assegurando, dessa forma, uma boa gestão dos dados”. (SANTOS, 2017, p. 124).

Enquanto os princípios FAIR caracterizam os dados científicos quanto à sua abertura, os *Panton Principles* caracterizam os dados científicos quanto ao seu licenciamento. A *Panton Principles*⁶ desenvolveu quatro recomendações para a abertura de dados científicos (MURRAY-RUST et al., 2010, não paginado, tradução nossa):

1. Quando os dados ou coleções de dados são publicados, é essencial que sejam publicados com uma declaração clara e explícita dos desejos e expectativas dos editores em relação à reutilização e redefinição de elementos de dados individuais, toda a coleta de dados e subconjuntos da coleção. Esta declaração deve ser precisa, irrevogável e baseada em uma declaração legal apropriada na forma de uma renúncia [tipo de licença CCZero];
2. Muitas licenças amplamente reconhecidas não se destinam a, e nem são apropriadas para dados ou coleções de dados. Existe uma variedade de isenções e licenças projetadas e apropriadas especificamente para o tratamento de dados. As licenças *Creative Commons* (além do CCZero), GFDL⁷, GPL,⁸ BSD⁹, etc. não são apropriadas para dados e seu uso é fortemente desencorajado;
3. O uso de licenças que limitam a reutilização comercial ou limitam a produção de obras derivadas, excluindo o uso para fins específicos ou por pessoas e organizações específicas, é fortemente desencorajado. Essas licenças tornam impossível integrar e redefinir efetivamente os conjuntos de dados e impedir atividades comerciais que possam ser usadas para oferecer suporte à preservação de dados; e
4. Além disso, na ciência, é fortemente recomendado que os dados, especialmente quando financiados com recursos públicos, sejam explicitamente colocados em domínio público por meio do uso da Licença de Dedicção do Domínio Público ou da Renúncia Zero [CCZero] à *Creative Commons*. Isso está de acordo com o financiamento de muitas pesquisas

⁶ Panton Principles. Disponível em: <https://pantonprinciples.org/index.html>. Acesso em: 16 set. 2019.

⁷ GNU Free Documentation License (GFDL).

⁸ General Public License (GPL).

⁹ Berkeley Software Distribution (BSD).

científicas e o *ethos* geral de compartilhamento e reutilização na comunidade científica.

Tanto os Princípios FAIR quanto os Panton Principles tornam clara que a abertura está diretamente relacionada ao tipo de licenciamento atribuído ao dado científico, indo além da sua disponibilização em repositórios e/ou outros ambientes. O tipo de licenciamento é o tónus que define as possibilidades de trabalho com os dados. Nesse aspecto, os *Panton Principles* são taxativos em sua recomendação para atribuição de licenças específicas para dados, pois nem todos os tipos de *Creative Commons* possuem abertura irrestrita para modificações, compartilhamento, reuso e usos comerciais, enquanto que nem todos os dados FAIR são necessariamente abertos para compartilhamento e reuso.

Nessa acepção, observa-se que a discussão sobre direitos de propriedade intelectual aplicada aos dados científicos é de extrema relevância e corrobora com os desafios sócio-cultural, organizacional e legal apontados por Bueno de La Fuente (2016) e a dimensão ético-legal-cultural observada por Oliveira e Silva (2016) quanto a disponibilização de dados científicos de pesquisa.

Assim, na lógica de abertura irrestrita para os dados científicos, Oliveira e Silva (2016, p. 25) ressaltam que “na cultura do compartilhamento, a visão é que o autor não é o proprietário absoluto de seus textos, escritos e obras, mas um compilador de idéias, textos, e agrupamentos de conhecimentos que geram um novo sentido e, por conseguinte, uma nova obra”. Ainda conforme as autoras, a abertura irrestrita no contexto da ciência aberta altera também a definição de autoria em si, pois a “autoria perde o status de propriedade e ganha o de colaboração”. Não obstante, se alguns dos benefícios do compartilhamento de dados científicos é a visibilidade e aumento de citação (PIWOWAR; DAY; FRIDSMA, 2007; TENOPIR et al., 2011; LAW, 2005; KIM; STANTON, 2015; CURTY, 2016; CURTY et al., 2017), logo, uma licença do tipo CCZero não acrescentaria ao pesquisador esses benefícios e, dessa forma, poderia inibir comportamentos de compartilhamento dos pesquisadores.

2.2.2 Gestão de dados científicos de pesquisa

Um aspecto importante a considerar na gestão de dados é o tempo de vida dos dados em uma investigação. Sayão e Sales (2015) publicaram o Guia de Gestão de Dados de Pesquisa, editado para bibliotecários e pesquisadores, que além de definir conceitos serve de manual para a construção de um Plano de Gestão de Dados (PGD), reconhecendo o potencial informacional dos dados de pesquisa para a ciência. Neste guia, os autores apresentam um modelo de referência sobre o ciclo de vida dos dados constituído em oito etapas, conforme abaixo:

FIGURA 4 - CICLO DE VIDA DOS DADOS



FONTE: Sayão e Sales (2015, p. 12).

A primeira etapa "Planejar" envolve a premeditação do projeto, no sentido de criação, delineamento metodológico, estabelecimento de infraestrutura tecnológica para todos os estágios dos dados, bem como todas as técnicas e sistemas que serão empregados no trabalho e suporte com os dados. É nesta etapa que será elaborado o PGD que servirá com instrumento orientador da pesquisa no manejo dos dados, do começo ao fim da investigação.

A etapa de "Coleta" pode compreender tanto a criação de dados quanto a obtenção destes em fontes externas. Nos dois casos, é necessária atenção às

condições “intrínsecas e extrínsecas relacionados à natureza e às condições de coleta” (OLIVEIRA, 2017, p. 175). A criação ou coleta dos dados envolve a infraestrutura tanto de tratamento quanto das condições de armazenamento durante a vida do projeto. Nesta etapa define-se também o ambiente de armazenamento dos dados de acordo com seu tipo (sensíveis, sigilosos, estáticos e dinâmicos), formato, linguagem, volume e padrões.

A terceira etapa “Assegurar a qualidade” está relacionada aos controles de qualidade e inspeção de dados, no sentido de garantir sua validação e acondicionamento. Para Oliveira (2017, p. 177), à esta etapa estão atrelados “a definição de procedimentos legais, éticos, normativos e critérios para as condições e restrições de uso (aberto, fechado, embarcado)”. Em suma, esta etapa garante a integridade do dado criado e/ou coletado.

Na etapa “Descrever” ocorre o processo de representação de um dado a partir de suas características, como por exemplo: a tipologia, formato, extensão, volume, projeção, autoria, edição/versão, proveniência, dentre outros. Cada característica de um dado é representada por um elemento chamado metadado. Por meio dos metadados é possível classificar, organizar e evidenciar a relação com outros objetos. Segue um breve resumo dos tipos de metadados no Quadro 1.

QUADRO 1 - TIPOS DE METADADOS

Tipo de metadado	Função	Exemplo	Padrão
Administrativo (e técnicos)	Provê informações sobre origem e manutenção de recursos; informações sobre <i>software</i> e <i>hardware</i> , formas de acesso;	Data de criação ou aquisição; direitos autorais; permissões e segurança; formato; tamanho de arquivo; compressão;	OAI; METS; Dublin Core;
Descritivo	Utilizado para descrição dos recursos para fins de descoberta e identificação (classificação e indexação);	Título; autor; resumo; descritores; contato;	HTML; XML; Dublin Core; TEI; EML; DOI; Handle System;
Estrutural	Informações que documentam a estrutura da fonte dos dados; é a organização de recursos complexos, das partes de um recurso entre si;	Acesso via WWW; Requisitos do sistema;	METS;
Preservação	Gestão a longo prazo de recursos digitais	Título original; versionamento; autenticidade; ambiente técnico; proveniência.	PREMIS (OAIS).

FONTE: Adaptado de Sayão (2010) e Sales (2014).

Os padrões de metadados são um conjunto de metadados que seguem regras específicas que evidenciam cada aspecto do dado. Os padrões são diversos, e entre eles, os mais conhecidos são: *Dublin Core* (DC), *International Standard Organization* (ISO), OAI-ORE, CERIF (ciências sociais, humanidades, genéricos); *Genome Metadata*, *Ecological Metadata Language* (EML), *Darwin Core*, ORE-XML (ciências biológicas, ambientais e agrícolas); *Directory Interchange Format* (DIF), ISO 19115 (ciências da Terra); CIF, AVM, NeXus (Física); entre outros. A adoção de padrões é primordial para a descoberta, preservação, intercâmbio, uso, compartilhamento e reuso de dados.

A quinta etapa é “Preservar”. Aqui, os dados são organizados e documentados para fins de armazenamento e preservação a longo prazo. Deve-se observar os padrões e protocolos apropriados que garantam a agregação, curadoria, interoperabilidade, formas de acesso e padrões de autoria abertos objetivando a distribuição.

“Descobrir” é a sexta etapa do ciclo de vida do dado científico e concentra-se nas possibilidades de compartilhamento e visibilidade dos dados por meio ambientes como repositórios, diretórios, metabuscadores, entre outros. A descoberta de dados está completamente vinculada à fase de descrição, pois uma vez descritos de forma apropriada e armazenados em ambientes seguros e íntegros, os dados são passíveis de serem acessados, compartilhados e reusados.

Embora o modelo do ciclo de gestão de dados de Sayão e Sales (2015) apresente a etapa “Integrar” como item distinto no ciclo, compreende-se aqui que a integração já faz parte da etapa “Descrever”, “Preservar” e “Descobrir”, pois estas estão interligadas a partir da definição de padrões de metadados e dos protocolos de intercâmbio dos repositórios. Em todo caso, a integração vale-se de protocolos de interoperabilidade como o *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE) ou o *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH).

A “Análise” é a oitava e última etapa do ciclo de gestão de dados. Nessa etapa, os dados já estão prontos para uso e passíveis de serem trabalhados por diversas ferramentas. Nesta etapa, o objetivo é garantir que o dado possa ser avaliado, remoldado, re combinado e reproduzido.

Todas as etapas do ciclo de gestão de dados buscam garantir organização e qualidade para a promoção do reuso. O conceito de reuso pode ser compreendido

“de maneira ampla como uso de dados – normalmente sem explícita permissão – para estudos previstos ou não pelo autor original dos dados ou por outros pesquisadores” (SAYÃO; SALES, 2014, p. 82).

Anterior à realidade nacional, em 1997, o *Inter-university Consortium for Political and Social Research* (ICPSR), parte integrante do *Institute for Social Research* dos Estados Unidos, lançava a primeira edição do seu guia para preparação e arquivamento de dados em ciências sociais com considerações sobre o compartilhamento de dados científicos e instruções para preparação, descrição (metadados), tratamento, armazenamento, aspectos de segurança, propriedade intelectual, questões éticas, preparação de coleções de dados, aspectos de qualidade, formatação, preparação para compartilhamento e depósito e elaboração de plano de gestão de dados.

O ciclo de gestão de dados também encontra representação dentro de outros ciclos, como o ciclo de vida da pesquisa (TENOPIR et al., 2011; HOFELICH MOHR; JOHNSTON; LINDSAY, 2016), regime de acesso de dados (ARZBERGER, et al., 2004) e ciclo de vida da curadoria de dados (HIGGINS, 2008), sendo a gestão dos dados um sub-processo dos demais ciclos. Nesse sentido, constata-se a importância de planejar a coleta, a descrição e a preservação dos dados de maneira a assegurar a sua qualidade para que outros pesquisadores possam reusá-los, descobrindo novos dados, integrando e analisando os dados já existentes (BORGMAN, 2012).

O ciclo de vida dos dados de pesquisa tem papel fundamental no planejamento de como os dados serão gerenciados no desenvolvimento do projeto e como serão compartilhados ao final. O PGD é um instrumento que engloba todas as etapas do ciclo de vida dos dados, uma vez que consiste em um documento formal que descreve quais dados serão coletados ou gerados pela pesquisa, bem como quais serão os métodos de coleta, tratamento, preservação e sua abertura à comunidade (SAYÃO; SALES, 2015).

Embora os PGD tenham se popularizado nos últimos 15 anos acompanhado o desenvolvimento dos primeiros guias e políticas de dados, em 1994 em artigo publicado na revista *BMJ*, o pesquisador em epidemiologia e saúde pública George Smith já alertava sobre a importância da abertura dos dados como forma de evitar fraudes em pesquisas. Para Smith (1994, p. 1519, tradução nossa), “os planos de gestão de dados podem se tornar um requisito para protocolos sustentando a visão

acessível de que os dados pagos pelo dinheiro público são de propriedade pública”. O pensamento de Smith tem relação direta com área do conhecimento de sua atuação. As ciências da saúde possuem forte comprometimento com a disponibilização de dados conforme atestam as pesquisas sobre comportamento em compartilhamento e reuso nesta área (PIWOWAR; DAY; FRIDSMA, 2007; TENOPIR et al., 2011; CURTY, 2015; GRIGG, 2015; JORGE; ALBAGLI, 2018).

O ciclo de vida do dado científico apresenta algumas semelhanças com o ciclo de gestão da informação de Choo (2002 citado por Araújo, 2011, p.41), pois compartilham processos distintos e inter-relacionados de necessidade (dado ou informação), aquisição, gerenciamento, tratamento, uso e reuso com a mesma finalidade, ainda que imputados com recursos diferentes, conforme quadro abaixo:

QUADRO 2 - SEMELHANÇAS ENTRE O CICLO DE VIDA DO DADO E O CICLO DE VIDA DA GESTÃO DA INFORMAÇÃO

Estágio / Processo	Ciclo de vida do dado	Processos de Gestão da Informação
	Pergunta de pesquisa	Necessidade de informação
	Pesquisa/reuso	Aquisição de informação
	Plano de Gestão de Dados	Organização
	Descrição	Tratamento
	Arquivamento	-
	Publicação	Uso da informação
	Reuso	Comportamento adaptativo

FONTE: Adaptado de Higgins (2008) e Choo (2002).

Nota-se que os estágios do ciclo de vida dos dados e os processos de gestão da informação são semelhantes devido à natureza da área do conhecimento de onde provém: a gestão da informação na Ciência da Informação. Se por um lado os processos de gestão informação já se encontram fortemente estabelecidos na literatura científica da área, os estágios do ciclo de vida do dado podem ser melhor compreendidos devido a essa semelhança, ainda que tratando de objetos diferentes.

Pode-se estabelecer então que dados científicos são dados primários e que podem ser “[...] armazenados em grandes centros de dados suportados por uma infraestrutura que possibilita a produção, armazenamento, localização, compartilhamento e reuso” (MEDEIROS; CAREGNATO, 2012, p. 318), sendo essa ciberinfraestrutura um suporte para a curadoria de dados realizadas em repositórios.

A compreensão do dado científico enquanto insumo e parte do resultado do processo de investigação indicam diversas possibilidades advindas do seu

compartilhamento e reuso para propiciar novas descobertas e inovações, o que será discutido na seção seguinte.

2.3 COMPARTILHAMENTO DE DADOS CIENTÍFICOS

Como definição básica, o compartilhamento de dados científicos pode ser definido como processo de compartilhar dados científicos de pesquisa para a utilização por outros. (BORGMAN, 2012). O compartilhamento está associado também à preservação, acesso e uso de dados, facilitando o desenvolvimento da ciência (PIWOWAR; DAY; FRIDSMA, 2007; TENOPIR et al., 2011).

O compartilhamento de dados é uma atitude que parece simplista à primeira vista, entretanto, existem diversas barreiras que impedem um pesquisador de compartilhar seus dados. Ainda de acordo com Borgman (2012), quanto mais difícil forem as fases de coleta e análise de dados, menor é a probabilidade de o pesquisador compartilhar seus dados. Esse comportamento pode ser explicado devido aos imbricados pormenores de todas as fases do ciclo de gestão de dados, principalmente durante os processos de coleta, codificação e organização dos dados.

Frederick (2016, p. 11) apresentou um resgate histórico sobre qual seria o “grande negócio” sobre o compartilhamento de dados e apresentou o dado como sendo o fundamento básico da ciência, o lugar de onde provém os resultados e assim, uma representação de eventos que já ocorreram, ocorrem e podem ocorrer. E nesse sentido, a autora comenta sobre a pesquisa de David (2000) que discorre sobre as origens da ciência aberta e que há uma série de razões para uma quase indisponibilidade dos dados científicos que reside em quatro aspectos:

(i) a supressão de dados é fortemente influenciada e moldada pelo capitalismo e consumismo - nesse sentido, o autor cita pesquisas que são desenvolvidas com recursos da iniciativa privada ou mesmo governamentais, e pesquisas que tem por objetivo a vantagem competitiva e/ou lucro, e dessa forma, seus dados não devem ser compartilhados e reusados para preservar o segredo industrial/comercial;

(ii) a natureza competitiva do mercado capitalista pode por vezes não revelar se determinado produto ou serviço é prejudicial às pessoas, animais ou meio

ambiente, ou a combinação dos três, o que poderia ser prejudicial à continuidade de financiamento;

(iii) o modelo de negócio dos periódicos que exigem pagamento mesmo que seja para cobrir os custos de publicação e o viés de publicação/citação que levam editores a preferencialmente publicarem resultados de pesquisas novas com resultados positivos e a quase inexistência de publicações cujos resultados foram negativos ou cuja hipótese fora refutada. Sobre isso, Frederick (2016, p.14) reflete que na realidade, o que é publicado é um “artigo revisado por pares e não dados revisados por pares” e que sendo os dados um assunto crítico para a ciência, é “desconcertante” que os dados sejam o menos importante em muitas pesquisas publicadas; e

(iv) a capacitação dos pesquisadores (ou a ausência dela) interfere na forma como os dados são tratados e isso impacta na disponibilização, compartilhamento e diretrizes para reuso.

Friderick fez ressalva nesse último aspecto e comenta que as universidades podem passar para os pesquisadores a impressão de que disponibilizar dados não é importante devido a ausência de repositórios e políticas mandatórias. A autora faz crítica também às plataformas de periódicos que não oferecem suporte para dados. Por fim, aponta que as bibliotecas têm papel relevante na gestão de dados, seja no tratamento, disponibilização e no letramento informacional sobre dados para a academia.

No contexto nacional, as pesquisas apontam para questões políticas ou de financiamento e não ao comportamento de pesquisadores quanto a compartilhamento e reuso.

No Brasil, as Fundações Estaduais de Amparo à Pesquisa (FAP's) são os órgãos responsáveis pela coordenação e articulação dos interesses das fundações de amparo à pesquisa, dos interesses da comunidade científica e das diretrizes governamentais na gestão de recursos públicos para fins de pesquisa (CONFAP, 2019). Nos últimos anos, as agências de fomento e demais fundações começaram a exigir dos pesquisadores o compromisso quanto a elaboração de um Plano de Gestão de Dados e a disponibilização em acesso aberto aos dados brutos utilizados tanto na elaboração do relatório (artigo, comunicação, livro, etc.) quanto os dados brutos enquanto produto da pesquisa.

Da forma prática, essa nova exigência das agências na concessão de financiamento provocou alterações na gestão dos processos de pesquisa pelo postulante ao financiamento. Agora, o pesquisador precisa planejar e detalhar todas as fases do projeto, descrevendo em formulários o roteiro preliminar completo, indicando onde e como serão levantados, tratados, convertidos e disponibilizados os dados científicos de pesquisa.

O Quadro 3 apresenta levantamento do contexto brasileiro quanto às políticas de gestão de dados requeridas pelas agências de financiamento, de forma a promover compreensão detalhada sobre a realidade local. Foi aplicada pesquisa documental nos sítios eletrônicos das fundações para verificar a existência de políticas de dados, recomendações ou indicação de repositórios para disponibilização dos dados. As políticas e/ou documentos de recomendação estão indicados no Apêndice 2.

QUADRO 3 - POLÍTICA DE GESTÃO DE DADOS REQUERIDAS PELAS AGÊNCIAS DE FOMENTO BRASILEIRAS

Agência/Entidade	Existência de Plano de Gestão de Dados?	Requer disponibilização de dados científicos?	Possui ou recomenda algum repositório? Qual?
CNPq	Não	Não	Não
FINEP	Não	Não	Não
CAPES	Não	Não	Não
EMBRAPA	Sim	Sim	Não
PETROBRAS	Não	Não	Não
FAPAC (Acre)	Não	Não	Não
FAPEAL (Alagoas)	Não	Não	Não
FAPEAP (Amapá)	Não	Não	Não
FAPEAM (Amazonas)	Não	Não	Não
FAPESB (Bahia)	Não	Não	Não
FUNCAP (Ceará)	Não	Não	Não
FADF (DF)	Não	Não	Não
FAPES (Espírito Santo)	Não	Não	Não
FAPEG (Goiás)	Não	Não	Não
FAPEMA (Maranhão)	Não	Não	Não
FAPEMAT (Mato Grosso)	Não	Não	Não
FUNDECT (Mato Grosso do Sul)	Não	Não	Não
FAPEMIG (Minas Gerais)	Não	Sim	Recomenda o repositório do Sistema de Informações sobre a Biodiversidade Brasileira (SiBBR)
FAPESPA (Pará)	Não	Não	Não
FAPESP (Paraíba)	Não	Não	Não
Fundação Araucária (Paraná)	Não	Não	Não
FACEPE (Pernambuco)	Não	Não	Não

FAPEPI (Piauí)	Não	Não	Não
FAPERJ (Rio de Janeiro)	Não	Não	Não
FAPERN (Rio Grande do Norte)	Não	Não	Não
FAPERO (Rondônia)	Não	Não	Não
Roraima – não tem fundação	-	-	-
FAPESC (Santa Catarina)	Não	Não	Não
FAPESP (São Paulo)	Sim	Sim	Recomenda repositórios externos e internacionais como Dryad.

FONTE: A autora (2019) com base nas informações dos sítios eletrônicos das fundações.

A análise do Quadro 3 indica que embora existam iniciativas nacionais, o Brasil ainda carece de uma política articulada de âmbito nacional para subsidiar, estruturar e organizar a informação científica de forma a prover tanto o acesso aberto à publicação científica e o compartilhamento de dados. De 28 fundações de amparo à pesquisa e/ou agências de fomento, apenas duas fundações requerem dos pesquisadores o PGD, a disponibilização aberta de dados científicos das pesquisas financiadas, e possuem ou recomendam repositórios abertos para disponibilização desses dados. Isto posto, demonstra a necessidade real de aprofundar as discussões sobre o compartilhamento dos dados científicos de pesquisa no Brasil, pois o número reduzido de ações concretas sobre compartilhamento oferece também uma visão sobre quão grande pode ser o potencial de exploração de dados.

O quadro acima se refere especificamente à cobertura para gestão de dados científicos e não teve pretensão de pontuar questões sobre acesso aberto de forma ampla nas citadas fundações. Os documentos analisados no Apêndice 2 fornecem pequenos indicativos de movimento de abertura, como por exemplo as chamadas públicas do CNPq que recomendam que a publicação dos resultados das pesquisas seja em periódicos de acesso aberto. Entretanto, não há indicação sobre a publicação dos dados científicos produzidos nas pesquisas. Entidades como EMBRAPA e FAPESP estão em fase de elaboração de seus repositórios de dados.

No Brasil, a liderança sobre a discussão acerca do compartilhamento de dados científicos foi iniciada pelo IBICT com o lançamento do "Manifesto de acesso aberto a dados da pesquisa brasileira para a ciência cidadã" (IBICT, 2016) que faz uma série de recomendações às universidades, institutos de pesquisa e pesquisadores sobre a criação de políticas institucionais mandatórias para assegurar a gestão de dados abertos, estimular a colaboração e a interdisciplinaridade,

estímulo ao desenvolvimento de infraestrutura tecnológica utilizando-se de sistemas abertos para a curadoria dos dados e interoperabilidade, principalmente.

A Rede Nacional de Ensino e Pesquisa (RNP) é outra entidade que está na vanguarda da discussão sobre a disponibilização, compartilhamento e reuso de dados e mantém ativo um grupo de discussão na *Research Data Alliance* (RDA)¹⁰ para apoiar as promoções da RDA no Brasil e fomentar o compartilhamento de dados com o estabelecimento de uma rede nacional e discussões sobre políticas públicas.

De maneira concreta, as universidades e institutos de pesquisa nacionais já mantêm ou estão em fase de elaboração dos seus repositórios de dados institucionais devido às próprias necessidades dos pesquisadores no trabalho colaborativo com outros atores e instituições, como por exemplo o consórcio criado pela FAPESP entre as universidades no estado de São Paulo¹¹, a Universidade Federal do Rio Grande do Sul (UFRGS),¹² Universidade Federal do Paraná (UFPR)¹³ e a Universidade de São Paulo (USP).¹⁴

No tocante a prática de compartilhamento, o relatório "Acesso aberto a dados de pesquisa no Brasil" (VANZ, et al., 2018, p. 22) revela que a prática de compartilhamento de dados no Brasil "não é algo tão comum" e baseia sua alegação no fato de que apenas 9,19% de 4.703 respondentes da pesquisa afirmaram que compartilham todos os seus dados, sendo que 37,02% não compartilham nenhum dado, conforme gráfico a seguir:

GRÁFICO 1 - PERCENTUAL DE DADOS COMPARTILHADOS EM ACESSO ABERTO

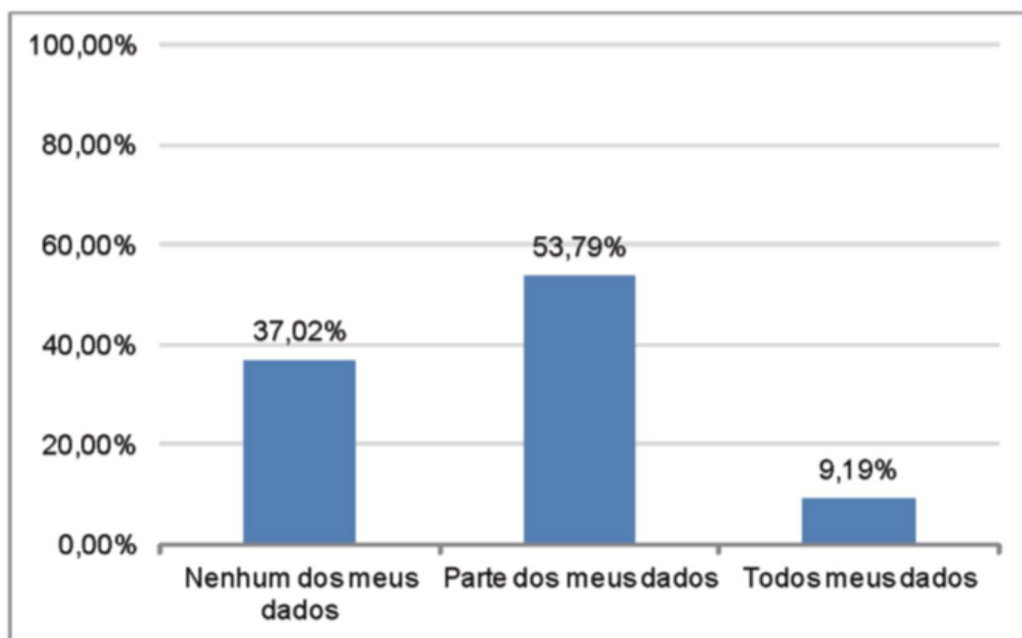
¹⁰ Disponível em: <https://www.rd-alliance.org/groups/rda-brazil>. Acesso em: 09 set. 2019.

¹¹ Disponível em: <http://metabuscador.sc.usp.br/>. Acesso em: 9 set. 2019.

¹² Disponível em: <https://cedap.ufrgs.br/jspui/handle/20.500.11959/90>. Acesso em: 09 set. 2019.

¹³ Disponível em: <https://bdc.c3sl.ufpr.br/>. Acesso em: 09 set. 2019.

¹⁴ Disponível em: <https://uspdigital.usp.br/repositorio/>. Acesso em 09 set. 2019.



FONTE: Vanz et al. (2019, p. 22).

Pesquisas internacionais sobre o compartilhamento de dados indicam que as políticas mandatórias das entidades reguladoras e financiadoras de pesquisa são e responsáveis em maior parte pelas ações de compartilhamento (GRIGG, 2015). Já a pesquisa de Tenopir et al. (2011, p. 9), a nível mundial, investigou o comportamento de compartilhamento e reuso de dados científicos e descobriu que áreas como astronomia, biologia, meio ambiente/ecologia e física são, respectivamente, as áreas que mais compartilham dados científicos. De modo inverso, áreas como medicina e ciências sociais estão entre as áreas que menos compartilham.

Nessa perspectiva, a identificação do impacto percebido pelos pesquisadores quanto ao reuso de dados científicos em repositórios brasileiros, objetivo geral desta pesquisa, pode somar-se a estudos na área e contribuir para a construção de um panorama brasileiro.

2.3.1 Vantagens do compartilhamento de dados

Ao visualizar dados como sendo a infraestrutura da pesquisa, uma boa gestão destes pode ser capaz de conferir vantagens aos processos de investigação. Numa realidade em que a coleta e produção de dados é exponencial e demanda quase que obrigatoriamente o trabalho colaborativo, o compartilhamento de dados é

cada dia mais relevante e apresenta inúmeras vantagens em sua adoção, os quais destacam-se (TENOPIR et al., 2011, p. 1):

- a reanálise de dados ajuda a verificar os resultados, o que faz parte do processo de validação científica;
- diferentes interpretações ou abordagens de dados contribuem para o progresso científico, especialmente na área interdisciplinar;
- uma boa gestão atrelada a uma preservação a longo prazo ajuda a manter a integridade de dados;
- a disponibilidade de dados permite a otimização na re-coleta de dados;
- a disponibilidade de dados oferece salvaguardas contra má conduta, falsificação e fabricação de dados; e
- estudos de replicação servem como ferramentas de treinamento para novas gerações de pesquisadores.

Além das vantagens destacadas, pesquisas sobre compartilhamento de dados relatam preocupações dos pesquisadores quanto a disponibilização de dados. Algumas dessas preocupações residem na definição do que é dado nas diversas áreas da ciência, falta de conhecimento/habilidades técnicas para compartilhamento, o esforço e custo requeridos na preparação dos dados para compartilhamento, preocupações com o *design* do ambiente dos repositórios que devem ser otimizados para performance em pesquisa científica e não somente para banco de dados, preocupações com uso secundário de dados e riscos à carreira de pesquisador [devido à interpretações equivocadas com as ideias do autor original] (BORGMAN, 2012; KIM; STANTON, 2015; FREDERICK, 2016).

Borgman (2012) apresentou pesquisa onde destacou quais são as preocupações dos pesquisadores quanto ao compartilhamento de dados. Em seu trabalho, a autora considerou as justificativas mais comuns encontradas na literatura sobre as vantagens do compartilhamento, os quais são: (i) reproduzir ou verificar pesquisas; (ii) disponibilizar ao público resultados de pesquisas financiadas publicamente; (iii) permitir que outras pessoas façam novas perguntas sobre os dados existentes; (iv) e avançar no estado de pesquisa e inovação levando em conta os possíveis beneficiários do compartilhamento e os argumentos para compartilhar. Como resultado, observou que o compartilhamento é um processo complexo e difícil, pois os pesquisadores têm grandes dúvidas sobre "que dado pode ser compartilhado, por quem, com quem, sob quais condições, o porquê e sob quais

efeitos". E, ainda, que a multiplicidade de formas e suportes sobre o que pode ser um dado dentro das diversas áreas da ciência leva a mais dúvidas sobre o que representa um dado ser interoperável e reutilizável, pois:

Interoperável pressupõe perícia suficiente para avaliar a integridade dos dados e entender seu significado. Isso pressupõe também documentação adequada do contexto de criação, processamento e proveniência de dados. Reutilizável é um padrão apenas aquém da reprodutibilidade. Especialização considerável, esforço, reestruturação e softwares proprietários podem ser necessários para reutilizar dados (BORGMAN, 2012, p. 1072, tradução nossa).

Assim sendo, observa-se que o compartilhamento de dados tem muitas nuances e envolve a abertura de processos complexos do ciclo de uma pesquisa científica que não é simples de tornar transparente. A condução de uma pesquisa antes da *e-Science* com vistas restritas somente ao seu grupo específico de trabalho percorria procedimentos formais sem ter a preocupação de traduzir esses caminhos a formatos abertos, plenamente documentados e legíveis. Os movimentos da ciência aberta e a pressão institucional ou de financiamento fazem com que esses caminhos científicos precisem ser readequados a uma lógica de abertura ainda formal, porém simplista e acessível para fins de reprodutibilidade - o que requer mais esforço, mais investimento e vontade.

2.4 REUSO DE DADOS CIENTÍFICOS

Em uma acepção básica, o reuso de dados pode ser definido com um dado compartilhado em algum ambiente e que fora utilizado por um outro alguém. Todavia, ainda que compartilhamento e reuso de dados estejam intimamente ligados, o que os diferenciam são os aspectos técnicos requeridos para a efetivação do reuso de dados (FANIEL; ZIMMERMANN, 2011). Esses aspectos técnicos são requisitos de “relevância, completude, compreensibilidade e a confiabilidade” do dado (CURTY, 2019, p. 179). Esses requisitos estão ligados à forma com o dado foi elaborado, analisado, codificado e organizado e essa estrutura é representada por meio da documentação de um dado.

O reuso de dados somente fará sentido quando os quatro requisitos técnicos acima citados encaixarem-se perfeitamente ao desenho de um projeto, sendo capaz de responder à pergunta de pesquisa.

A literatura aponta trabalhos que abordam as barreiras, dificuldades e riscos em promover o compartilhamento e reuso dos dados de pesquisas (GRINYER, 2009; TENOPIR et al., 2011; FANIEL; ZIMMERMAN, 2011; CURTY, 2016; CURTY et al., 2017; CURTY, 2019). No entanto, para além da elaboração dos dados, existem grandes desafios quanto a disponibilidade destes que vai desde a sua produção, passando pelo acesso até o reuso. De acordo com Bueno de La Fuente (2016, não paginado, tradução nossa) são elencados alguns desafios em relação a disponibilização de dados abertos de pesquisa:

- Sócio-cultural: estes vêm principalmente de pesquisadores e estão relacionados com a falta de consciência sobre os benefícios e importância da abertura de sua investigação; a relutância em mudar os seus fluxos de trabalho e as práticas atuais em relação à divulgação de dados ao longo do processo de investigação; pesquisadores o consideram como um tempo e atividade adicionando às suas cargas de trabalho existentes que consome esforço; as diversas abordagens que podem ter pesquisadores de diferentes disciplinas ou numa fase diferente em sua carreira ou a falta de um claro reconhecimento e um sistema de recompensa que promove práticas de ciência aberta.
- Tecnológico: embora as tecnologias de informação e comunicação atuais tenham evoluído, ainda há uma grande variedade de aspectos a melhorar, a fim de apoiar e facilitar os fluxos de trabalho dos pesquisadores para o trânsito para uma cultura de abertura.
- Política: há uma clara necessidade de um compromisso político para promover a ciência aberta e integrá-lo nas agendas governamentais. As novas políticas devem ser formuladas e desenvolvidas em estratégias claras. Além disso, os recursos necessários devem ser alocados para que essas políticas possam ter sucesso. Uma verdadeira ciência aberta só pode acontecer em nível internacional, assim, os diferentes ritmos e abordagens dos países irão influenciar a sua realização. Mas também dentro de instituições de pesquisa próprias há uma necessidade de desenvolvimento de políticas e planejamento estratégico.
- Organizacional: como a pesquisa geralmente ocorre dentro de um quadro institucional, a própria organização tem que estar pronta para facilitar a transição para uma cultura de investigação aberta, adequando suas unidades,

serviços e recursos humanos para acomodar um fluxo de trabalho em ciência aberta, portanto, desenvolver ou adquirir habilidades necessárias para tal.

- Econômica: mesmo que um dos principais argumentos para ciência aberta seja uma maior eficiência da investigação que, no longo prazo irá resultar em uma melhor utilização dos recursos nos sistemas científicos e de inovação, investimentos significativos têm que ser feitos no começo a fim de desenvolver o ecossistema técnico, político e organizativo da ciência aberta.
- Legal: um quadro de legislação clara deve ser desenvolvido em nível internacional que definam as regras para a divulgação de dados e outras entradas e saídas de pesquisa, ao mesmo tempo proteger direitos que não devem ser dispensados como a privacidade, as informações pessoais, interesses comerciais e segurança.

Os desafios acima são os mesmos enfrentados pelos movimentos de acesso aberto a informação, pois, de modo geral, são questões que estão relacionadas à dinâmica de trabalho dos pesquisadores, seus métodos e procedimentos de trabalho. Ainda que o aparato tecnológico da *e-Science* altere o ambiente e a forma como a informação científica tramita e é trocada entre os atores, os desafios permanecem os mesmos e apresentam ênfase no comportamento do pesquisador e do financiamento à pesquisa.

2.5 REPOSITÓRIO DE DADOS DE PESQUISA

A partir do movimento pelo acesso aberto às políticas de acesso, os repositórios institucionais alteraram a forma e principalmente o custo com que os resultados de pesquisa são disponibilizados. Com a constante intensidade em geração e uso dos dados, os repositórios de dados de pesquisa e as políticas de gestão dos dados têm dado o suporte para que os arquivos digitais de dados tenham um papel relevante, alterando a própria forma de construção da ciência. (RODRIGUES, 2008 citado por SAYÃO; SALES, 2012, p. 128-129).

Na gestão de dados científicos de pesquisa, o acesso e reuso dependem dos arquivos eletrônicos e seus formatos, da infraestrutura tecnológica para armazenamento, dos sistemas de recuperação da informação e visualização de dados, além das políticas públicas e institucionais que regulam todo o “desenvolvimento científico” (CALLON, 1993, p. 92).

Dessa forma, além dos sistemas tradicionais de buscas, os recursos abertos viabilizados pela retroalimentação da atividade científica ficam disponíveis ao pesquisador, ampliando seu escopo de fontes de informação de forma gratuita. O acesso aos repositórios institucionais que trazem os resultados de outras pesquisas e aos repositórios com acesso aos dados de outros pesquisadores permite a democratização da produção científica e em última instância, beneficia a sociedade como um todo.

Os repositórios de dados têm requisitos distintos dos repositórios institucionais e/ou temáticos para mediar o acesso aos dados. Sem as características inerentes aos repositórios de dados, os dados não podem ser acessíveis e, conseqüentemente, reutilizáveis. De acordo com Sayão (2017, p. 157):

A ampla base instalada de repositórios institucionais, que tem uma amplitude global, parece não ser perfeitamente adequada à gestão de dados de pesquisa em termos de concepção, modelo de dados e de fluxo de trabalho. Isso porque os repositórios institucionais foram projetados, desde sua origem, com o objetivo de arquivar e disponibilizar publicações acadêmicas; nos seus fluxos não há a preocupação em capturar e dar suporte às fases iniciais do trabalho de elaboração de produtos de pesquisa, sejam eles publicações ou coleções de dados. [...] é necessário a instalação de infraestruturas tecnológicas e gerenciais robustas que assegurem o depósito, arquivamento, preservação e acesso das coleções de dados de valor contínuo.

Para Corrêa (2016, p. 1.668), os repositórios desempenham papel fundamental na “preservação, integridade e disseminação de dados de pesquisa”. Neste sentido, o autor salienta que as atividades de pesquisa passarão a ter como centro de produção de conhecimento os repositórios de dados. E, nessa lógica, seguindo os princípios de dados científicos abertos, as instituições estão a desenvolver repositórios específicos para dados com serviços de curadoria para as atividades de gestão de dados. Assim, os repositórios de dados científicos caracterizam-se como “sistemas de informações técnicas e organizacionais que ajudam os pesquisadores a gerenciar, armazenar e fornecer seus próprios conjuntos de dados e a encontrar e acessar facilmente conjuntos de dados de outras fontes” (KINDLING et al., 2017, não paginado, tradução nossa).

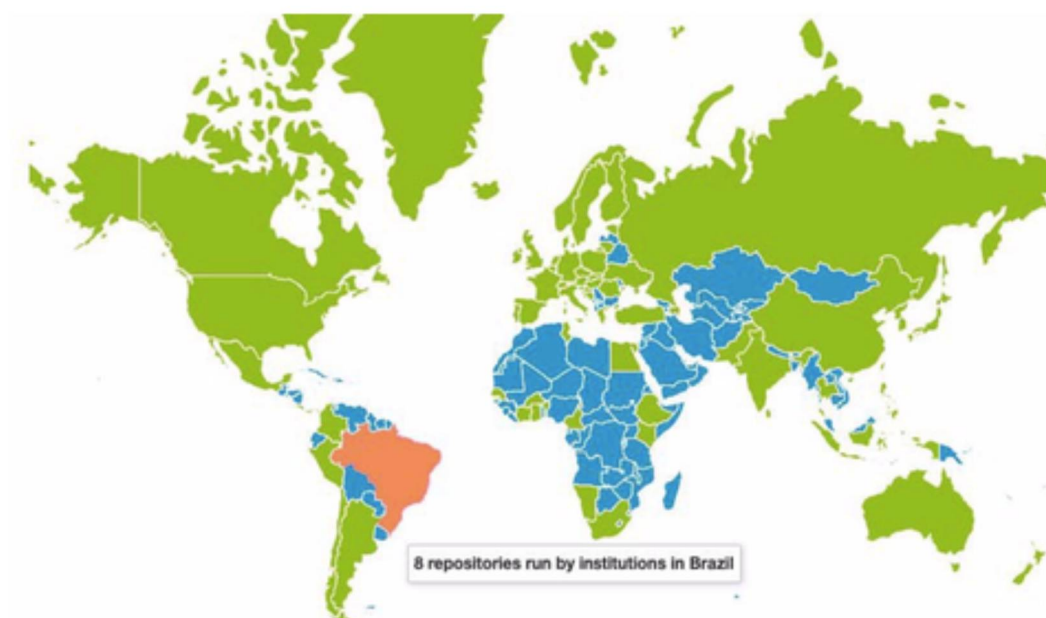
Inicialmente, essas estruturas eram voltadas para áreas que produziam grandes volumes de dados, como física, genética, astrofísica e ciências ambientais, exigindo infraestrutura de grandes proporções. Contudo, atualmente, a demanda por repositório de dados surge de um grande número de laboratórios menores ou

mesmo pesquisadores individuais que juntos produzem grande parte dos resultados científicos (SAYÃO, 2017).

Os repositórios podem armazenar e disponibilizar dados coletados por diferentes áreas de pesquisa. O Portal *Research Data Repositories Information* (Re3data.org) fundado em 2012 por uma parceria entre a Fundação Alemã de Pesquisa e a Universidade de Perdue nos Estados Unidos é um diretório internacional mantido desde 2016 pela organização global sem fins lucrativos DataCite¹⁵, e apresenta uma listagem dos principais repositórios de diversos países, detalhando tipos, contexto, conteúdo e características. O diretório pode ser utilizado como primeira opção na identificação de repositórios de dados científicos (PAVÃO; ROCHA; GABRIEL JUNIOR, 2018).

O Re3data.org exibe inicialmente a existência de 2.792 repositórios de dados registrados na plataforma¹⁶. Na aplicação de filtro para pesquisa de repositórios por país, verifica-se a existência de oito repositórios nacionais¹⁷.

FIGURA 5 - REPOSITÓRIOS BRASILEIROS REGISTRADOS NO RE3DATA



FONTE: Re3data.org. Pesquisa por país (2019).

¹⁵ Disponível em: <https://datacite.org/index.html>. Acesso em: 13 set. 2019.

¹⁶ Dados atualizados em 02 de out. 2019. Disponível em: <https://www.re3data.org/>.

¹⁷ Dados atualizados em 02 de out. 2019. Disponível em: <https://www.re3data.org/>.

O objetivo do Re3data.org é indexar e descrever informações detalhadas sobre os repositórios cadastrados por meio de 42 metadados abrangentes com as seguintes informações: a) metadados com informações gerais sobre os repositórios, como nome, URL, objetivo, área do conhecimento e descrição do serviço; b) informações sobre as instituições mantenedoras dos repositórios cadastrados, como nome, tipo do repositório, local, tipo de instituição e responsabilidade; c) informações relacionadas à questões legais, incluindo diretrizes, políticas de acesso e submissão de materiais; d) aspectos técnicos sobre identificadores persistentes para os conjuntos de dados dos repositórios, *software* utilizados e interfaces (RE3DATA, s/d).

De acordo com Vierkant et al. (2014, p. 4, tradução nossa),

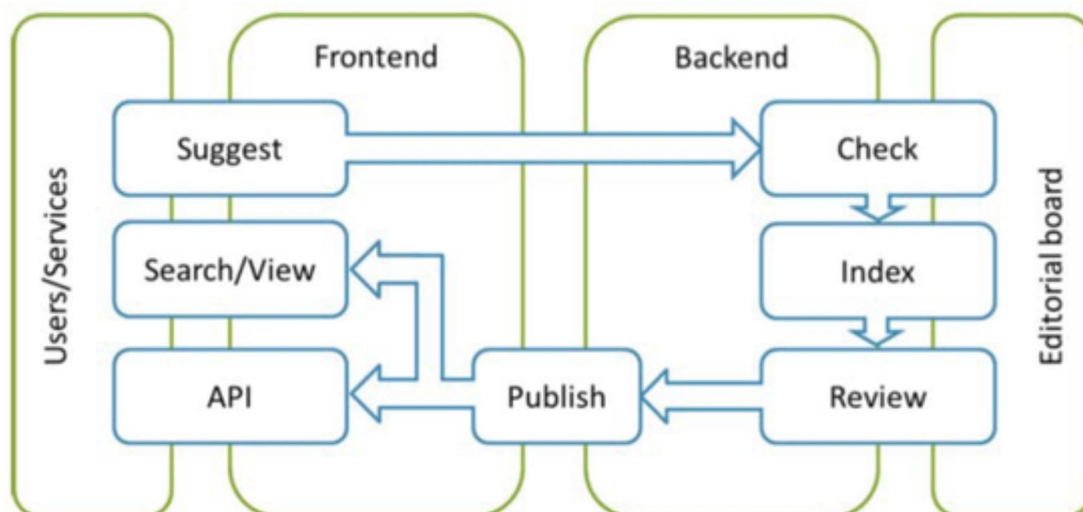
O esquema re3data.org é uma lista de propriedades de metadados que descrevem um repositório de dados de pesquisa sobre seu escopo geral, conteúdo e infraestrutura, bem como sua conformidade com padrões técnicos, de metadados e de qualidade. O esquema inclui propriedades de metadados necessárias e propriedades opcionais, fornecendo informações adicionais.

A estrutura do Re3data.org foi projetada para recomendar um padrão para descrição de repositórios de dados de forma a servir como base para a interoperabilidade entre repositórios e fornecer um padrão mínimo de requisitos para fins de certificação em repositórios de dados científicos de pesquisa (VIERKANT et al., 2014) e ainda “atuar como um provedor de dados que oferece principalmente seus dados de pesquisa e metadados ou um provedor de serviços que coleta dados de pesquisa de outros provedores de dados como base para criar serviços agregados” (RÜCKNAGEL et al., 2015, p. 4, tradução nossa).

O Re3data.org estabelece em sua política alguns critérios para registro de repositórios, os quais destacam-se abaixo:

- Foco em dados científicos de pesquisa;
- Ser administrado por uma entidade legal com estrutura organizacional que forneça sustentabilidade (exemplo: biblioteca, universidade, centros ou institutos de pesquisa);
- Tornar claras as condições de acesso ao repositório e aos dados; e
- Fornecer termos de uso.

FIGURA 6 - FLUXO DE REGISTRO DE REPOSITÓRIOS NO RE3DATA



FONTE: Rücknagel et al. (2015, p. 5).

Assim, o Re3data.org consolida-se como um diretório internacional para registros de repositórios de dados, além de oferecer metadados estruturais¹⁸ para repositórios e seguir como um padrão para a elaboração de repositórios.

Vierkant et al. (2014) e Rücknagel et al. (2015) recomendam que os repositórios tenham as seguintes funcionalidades:

- Suporte aos sistemas de identificadores persistentes para atribuir um DOI para cada *dataset* depositado no repositório;
- Licenciamento específico para dados a fim de tornar claras as condições de acesso e uso dos conjuntos de dados. Essas licenças devem também ser legíveis por máquina para garantir que a licença seja legalmente eficaz e compreensível globalmente;
- Disponibilização de metadados e *datasets* relacionados a outros serviços utilizados pelo repositório por meio de API como meio de melhorar a interoperabilidade dos serviços, a visibilidade de e a localização dos *datasets*;
- A adoção de padrões e certificados para garantir a confiabilidade e a confiança do repositório;

¹⁸ Esquema de metadados disponível em: <http://schema.re3data.org/3-0/re3data-template-V3-0.xml>. Acesso em: 02 out. 2019.

- Indicação de responsabilidade institucional ao usuário do repositório por meio da política de gestão de dados ou diretrizes;
- A criação de políticas e diretrizes para descrição dos serviços oferecidos pelo repositório e termos de uso; e
- A seleção de *software* para repositórios deve considerar um sistema de suporte padrões requeridos para o trabalho com dados.

Sob esse aspecto, a infraestrutura dos repositórios é determinante e pode facilitar o trabalho do pesquisador tanto nas atividades de depósito, compartilhamento e reconhecimento de autoria quanto no sentido de permitir o encontro de dados para fomentar pesquisas. Dessa forma, aspectos como confiabilidade, estabilidade e acessibilidade suportam a infraestrutura de repositórios (SAYÃO; LEITE, 2016).

Enquanto objetivo, “os repositórios devem garantir o acesso contínuo e aberto – agora e no futuro – aos resultados de pesquisa que se manifestam na forma de dados, e que são considerados parte importante do patrimônio digital da humanidade” (SAYÃO; LEITE, 2016, p. 96). Desta forma, é possível destacar que um repositório tem como perspectiva dar apoio à gestão dos dados científicos de pesquisa e, nesse sentido, conforme destacado por Sales (2014), os benefícios da adoção de repositórios de dados são:

- Promover a visibilidade dos dados;
- Promover o compartilhamento de dados;
- Conferir créditos ao autor dos dados;
- Suportar um ambiente para a preservação digital;
- Formação da memória científica e transparência;
- Segurança dos dados;
- Garantia de disponibilidade;
- Curadoria digital;
- Possibilidade de criação de serviços inovadores;
- Reuso de dados;
- Desenvolvimento de redes de repositórios por meio de protocolos de interoperabilidade; e
- Formação de indicadores de qualidade e produtividade institucionais.

As recomendações acima são princípios gerais de abertura para repositório de dados e não adentram detalhes técnicos e requisitos computacionais de infraestrutura.

As vantagens acima destacadas tornam claras que a adoção de repositórios de dados é uma função vital para a preservação, integridade e difusão de dados científicos uma vez que, por meio de uma rede de repositórios, torna-se possível a conexão entre comunidades por meio da inter-relação entre fontes de dados provenientes de distintas disciplinas em repositórios específicos ou multidisciplinares (SILVA, 2019).

Os repositórios de dados exclusivamente brasileiros que compõem o ambiente desta pesquisa serão oportunamente identificados e descritos na seção de apresentação e discussão de resultados.

2.6 O IMPACTO DOS DADOS CIENTÍFICOS NAS ATIVIDADES DE PESQUISA

As pesquisas sobre dados científicos comentam que o compartilhamento e reuso de dados gera impactos na atividade científica dos pesquisadores. Esse impacto é mencionado na forma de benefícios tais como: aumento de citações e créditos, visibilidade entre os pares, recompensas convites à colaboração em pesquisas conjuntas e economia de recursos na fase de coleta de dados, principalmente (ARZBERGUER et al., 2004; PIWOWAR; DAY; FRIDSMA, 2007; TENOPIR et al., 2011; KIM; BORGMAN, 2012; STANTON, 2015 e CURTY et al., 2017).

No entanto, a literatura da área situa frequentemente as dimensões de “impacto” dos dados dentro do contexto dos estudos métricos da informação, área que aplica técnicas específicas para medir e qualificar a produção científica por meio de indicadores capazes de aferir o comportamento de um campo científico (FREITAS; BUFREM; GRÁCIO, 2017).

Para Belter (2014) e Nielsen e Hjørland (2014), o uso de indicadores bibliométricos pode ser uma técnica eficaz para mensuração do impacto de dados, auxiliando pesquisadores e bibliotecários na elaboração e avaliação de índices de citação. Belter (2014) aponta que a problemática a cerca da falta de consenso sobre os elementos que compõem um padrão mínimo para citação dos dados influencia os índices de impacto. As diferentes entidades curadoras de dados elaboram

diferentes guias de citação, sendo o mais comum o guia da *Digital Curation Center* (DDC)¹⁹ que recomenda a presença de dez elementos mínimos para uma citação de dados (autor, data de publicação, título, edição, versão, URI, tipo de recurso, editora, UNF, identificador e localização).

Ball e Duke (2015) discorrem sobre impacto a partir de uma abordagem social-comportamental ampla, pois definem impacto como sendo a “[...] diferença que está fazendo dentro e fora da comunidade acadêmica” (p. 2) e “efeito ou influência que um evento, agente ou recurso exerce sobre o outro” (p. 3-4), relacionando conceitos como atenção (quantas pessoas sabem sobre o dado) e disseminação (quão amplamente um dado foi distribuído).

Para esses autores, a ênfase no impacto fornece evidências tangíveis de benefícios, pois

[a mensuração do impacto] significa ir além de uma análise bibliométrica tradicional de resultados acadêmicos para considerar como as necessidades mais amplas foram atendidas por esforços de pesquisas. [Os dados de pesquisa] podem ter impacto influenciando práticas ou políticas, gerando riqueza, impulsionando inovações industriais, enfrentando questões ou problemas sociais urgentes ou atendendo a uma necessidade urgente de uma comunidade (BALL; DUKE, 2015, p. 2, tradução nossa).

Ball e Duke (2015) citam o Índice-h e o Fator de Impacto (FI), da *Thomson Reuters* (atual *Clarivate Analytics*), calculados partir dos artigos indexados na *Web of Science* e publicados no *Journal Citation Report* (JCR) como medidas de contagem de impacto, sendo o primeiro para medir o impacto de publicações de um pesquisador e o segundo para medir impacto de periódicos científicos, respectivamente. Para os autores, esses índices fazem uma avaliação de impacto incompleta, pois essas métricas fazem avaliações a partir de um extrato de publicações e não da totalidade publicada. Ainda sim, os autores reconhecem que o impacto medido por meio de índices bibliométricos são medidas verdadeiras, pois os dados podem ser citados por razões de reconhecimento de influência e para cotação de financiamentos por agências de fomento.

Ferrer-Sapena et al. (2016) corroboram com Belter (2014) e Ball; Duke (2015) ao destacar que os indicadores bibliométricos podem ser um dos meios para medir o impacto de dados científicos de pesquisa, ainda que essas métricas dependam

¹⁹ Disponível em: <https://www.dcc.ac.uk/guidance/how-guides/cite-datasets#sec:elements>.

fundamentalmente da forma como é elaborada a referência do dado ou conjunto de dado citado. Ferrer-Sapena et al. (2016), assim como Ball e Duke (2015) recomendam o uso de diversos *rankings* fora do escopo do JCR para a medição de impacto, tal como *Altmetrics*, *Data Citation Index* e as métricas oferecidas por repositórios de dados gerais como *Figshare*, *Dryad* e *Open Aire*.

Dois estudos mostram uma abordagem comportamental do impacto dos dados científicos nos pesquisadores: Tenopir et al. (2012) e Curty et al. (2017).

A pesquisa de Tenopir et al. (2011) explorou as práticas e percepções dos pesquisadores em relação ao compartilhamento de dados e concluiu que os pesquisadores têm vontade de compartilhar os seus dados, ainda que percebam que são muitas as barreiras que impedem o compartilhamento e reuso efetivo. Essas barreiras consistem, principalmente, em obter créditos dos dados por meio referenciamento e citação formal dos *datasets*, a obtenção dos artigos derivados a partir do uso de seus dados, o estabelecimento de condições de uso e acesso aos seus dados e a adoção de padrões de metadados para garantir a qualidade de descrição e recuperação dos dados. Essas barreiras identificadas pelos autores podem dar uma dimensão do impacto negativo percebido pelos pesquisadores no compartilhamento e reuso de dados, pois as barreiras levantadas são condições que impedem tanto a disponibilidade quanto o reuso de dados.

O estudo de Curty et al. (2017) testou a relação entre as crenças e as atitudes dos cientistas em relação ao reuso de dados, bem como seu comportamento perante o reuso. Os resultados indicaram que a “aceitabilidade do reuso pode ser moldada pelo reconhecimento de uma boa reutilização” (p. 17, tradução nossa), ou seja, um comportamento de reuso pode ser adotado quando um pesquisador percebe os benefícios derivados da prática de reuso de dados. Em outras palavras, quando um pesquisador percebe a eficácia e eficiência advindas do reuso, pode-se prever que seu comportamento será de reutilizador de dados de outros cientistas/pesquisadores. Nesse sentido, o estudo de Curty et al. (2017) demonstra um impacto positivo do reuso de dados científicos, servindo de contraste ao estudo de Tenopir et al. (2011).

A partir dos autores acima, pode-se perceber que o impacto dos dados científicos é pensado em grande medida do ponto de vista da quantificação, enquanto o aspecto comportamental do impacto ainda é tratado de forma sutil e incipiente.

3 METODOLOGIA

Esta seção apresenta a metodologia utilizada na pesquisa, caracterizando-a do ponto de vista dos objetivos, procedimentos técnicos, o ambiente e amostra, a coleta de dados, instrumento de coleta e os procedimentos para análise dos dados.

3.1 CARACTERIZAÇÃO DA PESQUISA

A pesquisa é de natureza aplicada, com abordagem mista em relação ao problema de pesquisa. Conforme Gil (2019, p. 147), a abordagem mista “também pode ser utilizada quando o pesquisador deseja definir grupos de acordo com os resultados quantitativos e fazer o seu acompanhamento mediante pesquisa qualitativa”.

A considerar que a pesquisa objetiva buscar conhecer os sentimentos e pontos de vista que podem alterar a maneira como os pesquisadores se comportam quanto ao reuso de dados, a abordagem qualitativa na análise dos resultados mostra-se adequada, pois a “pesquisa qualitativa é usada para obter *insights* sobre os sentimentos e pensamentos das pessoas, o que pode fornecer a base para um futuro estudo qualitativo independente ou pode ajudar os pesquisadores a mapear instrumentos de pesquisa para uso em um estudo quantitativo” (SUTTON; AUSTIN, 2015, p. 226).

Do ponto de vista dos objetivos é do tipo exploratória e descritiva. Exploratória durante a coleta e tratamento de dados, no sentido de ampliar as percepções sobre impacto de pesquisa, e descritiva no propósito de descrever os fatores de impactos percebidos pelos pesquisadores quanto ao reuso de dados científicos (GIL, 2002).

Em relação aos procedimentos técnicos, utiliza-se da pesquisa bibliográfica para cumprimento do objetivo específico “a” e pesquisa documental para atingir os objetivos “b” e “c”. Para atender ao objetivo “d” aplicou-se um questionário eletrônico com questões fechadas. A adesão ao questionário justifica-se quando “quer-se ouvir um grande número de respondentes; estes estão em regiões geograficamente dispersas; tem-se um tempo mais restrito para a coleta de dados do que aquele necessário para fazer a entrevista” (VERGARA, 2012, p. 40). Perguntas fechadas foram utilizadas para identificar e posteriormente analisar as respostas recebidas.

Para a elaboração do referencial teórico e parcial alcance do objetivo específico “a”, foi realizada pesquisa bibliográfica nas bases de dados *Emerald*, *Scopus*, *SciElo*, *Science Direct* e *Web of Science* (coleção principal) aplicando às bases os seguintes termos em inglês e português: *open science*, *research data*, *scientific data*, *open scientific data*, *data repository*, *data reuse*, *data sharing* e *research data - impact*. A tipologia de material consistiu em artigos de periódicos, livros, capítulos de livros, teses e dissertações, sítios eletrônicos de organizações internacionais e guias específicos sobre dados científicos. Os materiais recuperados foram exportados para o gerenciador bibliográfico Mendeley a fim de excluir as duplicidades de arquivos e sistematizar as fontes de informação.

3.2 AMBIENTE E AMOSTRA DA PESQUISA

O ambiente da pesquisa são os repositórios brasileiros de dados científicos registrados no Re3data.org. A seleção de repositórios de dados foi realizada por meio de pesquisa documental (MARCONI; LAKATOS, 2017; GIL, 2008) envolvendo a consulta ao sítio do Re3data.org²⁰ que é um diretório global de registro de repositório de dados científicos (PAMPEL et al., 2013; VIERKANT et al., 2014; KINDLING et al., 2017). A justificativa para utilização do Re3data.org como ambiente é sua parceria e colaboração com diversos órgãos e entidades internacionais que trabalham para a promoção da ciência aberta e reutilização de dados científicos. Na análise, os repositórios listados no diretório como brasileiros estão descritos no Quadro 7.

O Quadro 7 apresenta a descrição preliminar dos repositórios que formam o ambiente da pesquisa, destacando informações de identificação, entidade mantenedora, área do conhecimento de atuação do repositório, povoamento (a quantidade de *datasets* depositados), quantidade de pesquisadores únicos cadastrados e tipo de licenciamento concedido aos dados. As informações foram coletadas via acesso ao sítio eletrônico dos repositórios indicados na ficha de resumo no Re3data.org.

²⁰ A lista de repositórios brasileiros está disponível em: <https://www.re3data.org>. Acesso em: 22 set. 2019.

QUADRO 4 - REPOSITÓRIO DE DADOS BRASILEIROS REGISTRADOS NO RE3DATA.ORG

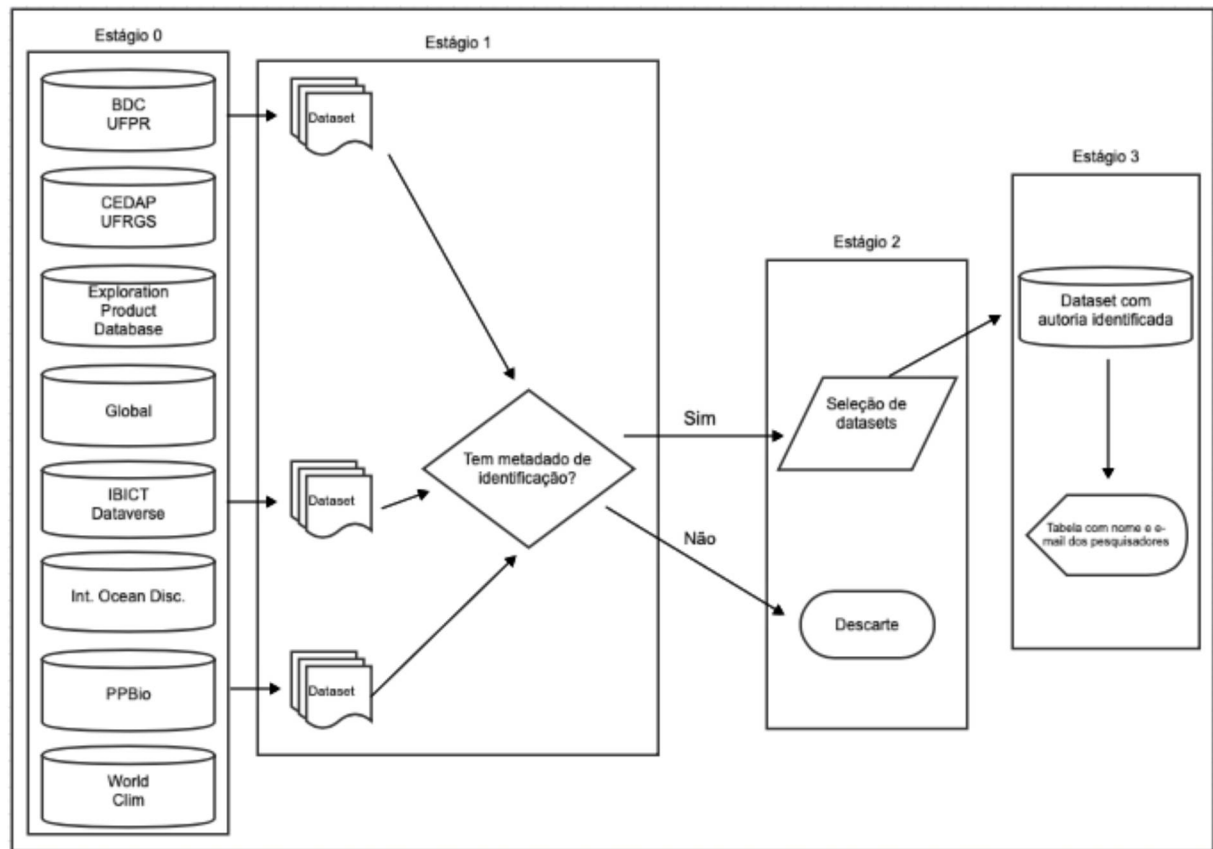
Repositório	Instituição mantenedora	Área do conhecimento	Qtde pesquisadores	Qtde datasets	Licenciamento dos dados
Base de Dados Científicos da UFPR (BDC/UFPR)	Universidade Federal do Paraná	Multidisciplinar	57	30	CC0; CC BY; CC BY-NC-ND
CEDAP <i>Data Research Repository</i>	Universidade Federal do Rio Grande do Sul	Multidisciplinar	Não identificado	25	Copyright
<i>Exploration and Production Database</i>	Agência Nacional do Petróleo e Gás (ANP) e Serviço Geológico do Brasil	Geofísica; Geologia	Não identificado	46	Não há informação sobre licenciamento
<i>Global Collaboration Engine (Globe)</i>	Consórcio Internacional; Representante brasileiro: INPE	Biologia; Botânica; Oceanografia; Geociências; Ciências naturais;	Não identificado	1	Copyright
IBICT <i>Dataverse Network</i>	Instituto Brasileiro em Ciência e Tecnologia	Multidisciplinar	Não identificado	240	CC0; restritos
<i>International Ocean Discovery Program</i>	Consórcio internacional; Representante brasileiro: CAPES	Oceanografia; Geociências; Geofísica; Geodésica; Geologia; Paleontologia; Ciências Naturais	Não identificado	22.068	Sem licenciamento definido. Mescla dados abertos e fechados
PPBio <i>Data Repository</i>	Centro de Estudos Integrados da Biodiversidade Amazônica	Biodiversidade; Ecologia; Botânica; Zoologia; Pesquisa de ecossistemas;	1.173	535	CC-BY
<i>WorldClim Global Climate Data</i>	Consórcio internacional; Representante brasileiro: Centro de Referência em Informação Ambiental (CRA)	Oceanografia; Geociências; Botânica.	Não identificado	32	Não permite compartilhamento

FONTE: A autora com dados do Re3data.org (2019) atualizados até 29 set. 2019.

Para o levantamento dos pesquisadores foi criado um fluxo de identificação em quatro estágios indicado na Figura 7. O estágio 0 consistiu na coleta dos *datasets* de cada repositório, que foram armazenados em banco de dados específico. O estágio 1 verificou a presença de metadados de autoria pessoal na descrição dos *datasets* (FIGURA 8). O estágio 2 selecionou os *datasets* que continham metadados de autoria. Estes foram marcados e arquivados em tabela específica. Os *datasets* que não continham autoria foram descartados. A etapa 3

consistiu na extração dos dados nome pessoal e *e-mail* dos pesquisadores, conforme seleção em tabela no estágio 2.

FIGURA 7 - FLUXO DE IDENTIFICAÇÃO DE PESQUISADORES



FONTE: A autora (2019).

Os repositórios CEDAP *Data Research Repository*, *Exploration and Production Database*, *Global Collaboration Engine (Globe)*, *International Ocean Discovery Program* e *WorldClim – Global Climate Data* foram descartados por não apresentarem metadados de autoria nos *datasets*, e também por esta razão não foi possível identificar a quantidade de autores/pesquisadores.

A identificação dos pesquisadores através do levantamento chegou ao quantitativo de 1.330 pesquisadores únicos. Todavia, a grande maioria dos pesquisadores não tinha *e-mail* cadastrado, o que reduziu consideravelmente o total para 336 pesquisadores.

FIGURA 8 - METADADO AUTORIA E ENDEREÇO DE E-MAIL

```

<?xml version="1.0"?>
<eml:eml
  packageId="doi:10.xxxx/eml.1.1" system="https://doi.org"
  xmlns:eml="https://eml.ecoinformatics.org/eml-2.2.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:stmml="http://www.xml-cml.org/schema/stmml-1.1"
  xsi:schemaLocation="https://eml.ecoinformatics.org/eml-2.2.0 xsd/eml.xsd">

  <dataset>
    <title>Primary production of algal species from Southeast Alaska, 1990-2002</ti
    <creator id="https://orcid.org/0000-0003-0077-4738">
      <individualName>
        <givenName>Matthew</givenName>
        <givenName>B.</givenName>
        <surName>Jones</surName>
      </individualName>
      <electronicMailAddress>jones@nceas.ucsb.edu</electronicMailAddress>
      <userId directory="https://orcid.org">https://orcid.org/0000-0003-0077-4738
    </creator>
    <keywordSet>
      <keyword>biomass</keyword>
      <keyword>productivity</keyword>
    </keywordSet>
    <contact>
      <references>https://orcid.org/0000-0003-0077-4738</references>
    </contact>
  </dataset>
</eml:eml>

```

FONTE: PPBio Data Repository; Jones et al. (2019).

De oito repositórios cadastrados no Re3data.org como brasileiros, apenas quatro são integralmente administrados por instituições brasileiras:

- Base de Dados Científicos da UFPR, vinculada à Universidade Federal do Paraná;
- CEDAP *Data Research Repository*, vinculado à Universidade Federal do Rio Grande do Sul;
- *Exploration and Production Database*, vinculada à Agência Nacional do Petróleo e Gás e Serviço Geológico do Brasil; e
- IBICT *Dataverse Network*, pertencente ao Instituto Brasileiro em Ciência e Tecnologia.

Os demais repositórios funcionam como consórcios interinstitucionais compostos por organizações sociais, universidades e institutos de pesquisa governamentais, não sendo exclusivamente nacionais.

Deste modo, o recorte da análise compreende apenas os repositórios exclusivamente brasileiros cadastrados no R3data.org. O recorte justifica-se para se ter a precisa identificação dos fatores de impacto sobre reuso dentro da realidade brasileira, pois a ausência de estudos sobre comportamento de reuso no Brasil foi a lacuna identificada na literatura durante o levantamento bibliográfico (TENOPIR et al., 2011; BORGMAN, 2012; CURTY, 2016; CURTY et al., 2017).

A considerar a definição de dados abertos pela *Panton Principles* elaborada por Murray-Rust et al. (2010, não paginado, tradução nossa) de que:

Por dados abertos na ciência, queremos dizer que ele está disponível gratuitamente na Internet pública, permitindo que qualquer usuário baixe, copie, análise, reprocessse, passe-os para o software ou use-os para qualquer outra finalidade, sem barreiras financeiras, legais ou técnicas do que aqueles inseparáveis de obter acesso à própria Internet.

E ainda a definição de Silva (2019, p. 33) a partir do qual “Os dados abertos são os que podem ser utilizados e distribuídos por qualquer pessoa, sem barreiras técnicas ou legais, sob os requerimentos de reconhecer a autoria e compartilhar o novo produto nas mesmas condições”, verifica-se que apenas três repositórios satisfazem as condições de abertura requeridas para processos de compartilhamento e reuso de dados, saber: Base de Dados Científicos da UFPR, IBICT *Dataverse Network* e PPBio *Data Repository*.

Os repositórios CEDAP *Data Research Repository* e *Globe* apresentam informação de que os dados são protegidos por *Copyright*; no repositório *Exploration and Production Database* os dados são fechados e disponibilizados mediante autorização para empresas com participação em processo licitatório para áreas de exploração de gás e petróleo; o repositório *WorldClim Global Climate Data* não permite compartilhamento de dados; e o repositório *International Ocean Discovery Program* não apresenta nenhuma informação sobre o licenciamento de dados, conforme Tabela 1.

TABELA 1 - REPOSITÓRIOS EXCLUÍDOS DA ANÁLISE

Repositórios Excluídos	Motivo de Exclusão
<i>Exploration and Production Database</i>	Repositório é de acesso restrito
<i>CEDAP Data Research Repository</i>	Dados são protegidos por Copyright
<i>Global Collaboration Engine (Globe)</i>	Dados são protegidos por Copyright
<i>International Ocean Discovery Program</i>	Não apresenta informação sobre licenciamento de dados
<i>WorldClim Global Climate Data</i>	Não permite compartilhamento de dados

FONTE: A autora (2019) com base nos repositórios selecionados.

Logo, os repositórios que foram ambiente para análise são apresentados na tabela abaixo:

TABELA 2 - REPOSITÓRIOS SELECIONADOS

Repositório	Quantidade Datasets	Pesquisadores Cadastrados
<i>PPBio Data Repository</i>	535	281
<i>IBICT Dataverse Network</i>	153	5
Base de Dados Científicos da UFPR	30	50

FONTE: A autora (2019) com base nos repositórios selecionados.

A amostra da pesquisa é do tipo não probabilística por julgamento. A pesquisa concentra-se nos 336 pesquisadores que disponibilizaram *datasets* nos repositórios exclusivamente nacionais registrados no Re3data.org. Esses pesquisadores estão concentrados em sua maioria na área de Biodiversidade e sistemas ecológicos e os demais na área multidisciplinar. Uma observação preliminar destaca que a maioria dos *datasets* possui múltipla autoria, um clássico indicador de colaboração científica.

3.3 COLETA DE DADOS

A coleta de dados ocorreu em duas fases. Na fase 1 foi aplicada a pesquisa documental para a identificação e coleta dos dados dos repositórios de dados no Re3data.org e as diretrizes e políticas dos respectivos repositórios. Ainda nesta fase foram coletados os dados dos pesquisadores com dados depositados nos repositórios e realizada a identificação e seleção dos *e-mails* dos pesquisadores para aplicação do questionário.

A fase 2 da coleta de dados compreendeu a aplicação do questionário para identificação dos pesquisadores e dados para análise dos perfis de comportamento sobre reuso de dados. As duas fases da coleta de dados estão descritas no quadro abaixo.

QUADRO 5 - FASES DA COLETA DE DADOS

Objetivos		Procedimentos	Coleta	Análise
Objetivo “a”	Identificar os fatores de impacto sobre reuso dos dados na literatura	Pesquisa bibliográfica	Levantamento bibliográfico	Referencial teórico
Objetivo “b”	Identificar e descrever os repositórios de dados brasileiros registrados no Re3data	Pesquisa documental	Levantamento dos repositórios no Re3data.org; Seleção dos repositórios exclusivamente brasileiros; e Descrição dos repositórios.	-
Objetivo “c”	Identificar nos repositórios brasileiros no Re3data os pesquisadores que apresentam <i>datasets</i> depositados	Pesquisa documental	Identificação dos pesquisadores; Seleção dos pesquisadores com e-mail cadastrados.	-
Objetivo “d”	Categorizar o impacto percebido pelos pesquisadores no reuso de dados científicos	Questionário fechado	Identificação dos fatores de impacto; Categorização dos fatores de impacto.	Estatística descritiva

FONTE: A autora (2020).

A considerar o objetivo específico “d” em categorizar os fatores de impacto sobre reuso de dados para uma amostra de 336 pesquisadores, o questionário foi o instrumento de coleta de dados selecionado para a obtenção do objetivo, pois o uso deste possibilita longo alcance geográfico, e atinge um grande número de pessoas, proporciona flexibilidade temporal para os respondentes, bem como a padronização de respostas (GIL, 2008; VIEIRA, 2009).

3.3.1 Elaboração do instrumento de coleta de dados

O instrumento de coleta de dados foi composto pelos seguintes elementos: carta de apresentação expondo os objetivos da pesquisa, instruções para

preenchimento, prazos, identificação e contato dos responsáveis; Termo de Consentimento Livre e Esclarecido (TCLE); e questionário.

O questionário foi elaborado observando o referencial teórico e os objetivos da pesquisa (QUADRO 6) projetado com 25 perguntas divididas em três partes, sendo:

- (i) caracterização dos pesquisadores com sete questões;
- (ii) comportamento dos pesquisadores quanto aos dados científicos de pesquisa com seis questões;
- (iii) comportamento dos pesquisadores quanto ao reuso de dados científicos de pesquisa com 12 questões.

O formato das questões foi padronizado com diferentes opções, tais como: do tipo binária e escalonada. As questões do tipo escalonadas utilizaram a escala de Likert com cinco pontos a fim de atribuir valores quantitativos para respostas qualitativas, sendo: (1) discordo fortemente; (2) discordo; (3) não concordo, nem discordo; (4) concordo; e (5) concordo fortemente.

O instrumento foi criado na plataforma *Google Docs* com o aplicativo *Forms* e pode ser visto no Apêndice 1.

QUADRO 6 - OBJETIVOS ESPECÍFICOS, RESULTADOS ESPERADOS, QUESTÕES E SUPORTE TEÓRICO

Objetivo	Resultados esperados	Questões	Suporte teórico
a) identificar os fatores de impacto sobre reuso de dados científicos na literatura;	<ul style="list-style-type: none"> • Verificar o impacto que os pesquisadores percebem em sua atividade de pesquisa ao reusar dados científicos; 	Questões 11 a 14; 18 a 25	ARZBERGUER et al. (2004); BORGMAN (2012); SAYÃO; SALES (2014); GRIGG (2015); TENOPIR et al. (2015); FREDERICK (2016); OLIVEIRA; SILVA (2016); CURTY et al. (2017); CURTY (2019); SILVA (2019).
b) descrever os repositórios de dados brasileiros registrados no Re3data;	<ul style="list-style-type: none"> • Verificar quais são os repositórios exclusivamente brasileiros cadastrados no Re3data; • Verificar se os repositórios são exclusivamente de dados ou se são repositórios digitais genéricos e/ou temáticos; • Verificar a presença de políticas e diretrizes de 	Questões 8 a 10; 15;	ARZBERGUER et al. (2004); RODRIGUES et al. (2010); PAMPEL et al. (2013); VIERKANT et al. (2014); SAYÃO; SALES (2016); COX et al. (2017); WU et al. (2019); LATIF; LIMANI; TOCHTERMANN (2019).

	gestão de dados nesses repositórios;		
c) identificar nos repositórios de dados brasileiros no Re3data os pesquisadores que apresentam <i>datasets</i> depositados; e	<ul style="list-style-type: none"> Conhecer o perfil dos pesquisadores que compartilham e reusam dados científicos de pesquisa. 	Questões 1 a 8; 16 e 17	PAMPEL et al. (2013); SAYÃO; SALES (2014); MOHR et al. (2016).
d) categorizar o impacto percebido pelos pesquisadores no reuso de dados científicos.	<ul style="list-style-type: none"> Identificar, com base das respostas dos pesquisadores e com auxílio do referencial teóricos, as categorias de impacto percebidas pelos pesquisadores quanto ao reuso de dados científicos; Verificar são as categorias que mais impactam em suas atividades científicas. 	Questões 14; 18 a 25	FISHBEIN; AJZEN (1975); SMITH (1994); ARZBERGUER et al. (2004); LAW (2005); VILLAVECES et al. (2005); PIWOWAR; DAY; FRIDSMA (2007); MOUTINHO; ROASSI (2010); TENOPIR et al. (2011); BORGMAN (2012); MEDEIROS; CAREGNATO (2012); BELTER (2014); NIELSEN; HJØRLAND, 2014; BALL; DUKE, 2015; GRIGG (2015); KIM; STANTON (2015); FERRER-SAPENA et al. 2016; SUTTON; AUSTIN (2015); CURTY et al. (2017).

FONTE: A autora (2019).

3.3.2 Aplicação do pré-teste

Uma versão preliminar do questionário foi apresentada para um grupo de quatro pesquisadores acadêmicos da área de dados abertos. A reunião ocorreu nos primeiros dias de novembro de 2019 e foram apontadas as adequações abaixo:

1. redação da questão 11 resposta nº 2;
2. reordenamento das questões 11 e 12; e
3. alteração no *layout*.

Após adequação do instrumento, o pré-teste foi submetido no período de 17 a 23 de dezembro de 2019 para uma amostra de dez pesquisadores com características semelhantes às da amostra. Os indivíduos que realizaram o pré-teste não fizeram parte da amostra selecionada para a aplicação efetiva do questionário.

O pré-teste obteve como resultado sete respostas, nas quais não foram relatados constrangimentos na abordagem ou outras dificuldades. A redação das questões 17 e 21 foram alteradas.

3.3.3 Procedimento de coleta e análise dos dados

A coleta de dados na fase 2 foi realizada através do envio do questionário pela plataforma de correio eletrônico Gmail para os 336 pesquisadores identificados no levantamento. O primeiro envio foi realizado no dia 09 de janeiro de 2020. O questionário aceitou respostas até 10 de fevereiro de 2020.

O primeiro envio apresentou retorno automático de 55 *e-mails* que “não puderam ser entregues porque o endereço não foi encontrado ou não pode receber mensagens”. Foi recebido *e-mail* de uma pesquisadora a solicitar esclarecimento adicional acerca da questão nº 19, o qual foi prontamente respondido.

Os *e-mails* retornados foram verificados para checagem de erros de grafia ou digitação. Após esse procedimento, foram novamente encaminhados para os pesquisadores, sendo que 52 *e-mails* retornaram como indisponíveis. Um pesquisador respondeu agradecendo o envio questionário e solicitando compartilhamento dos resultados ao fim da pesquisa.

No seu fechamento, o questionário obteve 73 respostas, representando um retorno 21,72%. A taxa de retorno pode ser compreendida devido a natureza flexível dos questionários eletrônicos de aplicação remota, pois não há controle sobre a entrega dos respondentes. Para Freitas, Janissek-Muniz e Mascarola (2004, p. 7), “um dos problemas frequentes [com questionários *on-line*] é que diversos endereços eletrônicos atribuídos não estão mais ativos”. Outra razão para explicar a taxa de retorno pode ser a correspondência entre a data de envio e o período de início de ano compreendido como férias de docentes.

As respostas do questionário foram recebidas em planilha do Google Forms e em seguida foram exportadas para o *Excel* para organização, tabulação e codificação dos dados. Os dados foram submetidos ao *software* SPSS Statistics para cálculo de estatísticas descritivas a fim de verificar a distribuição de frequência da amostra e posterior interpretação dos resultados obtidos.

A fim de avaliar a confiabilidade das respostas, foi aplicado o Alfa de Cronbach que, de acordo com Hair et al. (2009) e Matthiesen (2011), deve receber valores iguais ou superiores a 0,70 para garantia de respostas satisfatórias. O valor do Alfa foi de 0,712 atestando a confiabilidade das respostas (TABELA 3). O cálculo completo do Alfa pode ser verificado no Apêndice 3.

TABELA 3 - COEFICIENTE ALFA DE CRONBACH

Alfa de Cronbach	Nº de Itens
0,712	101

FONTE: Dados da pesquisa (2020).

Atestada a consistência interna das respostas, ato seguinte foi analisá-las utilizando as medidas de dispersão Desvio Padrão, Assimetria e Curtose.

A análise descritiva dos dados foi segmentada pelas três partes do questionário, conforme os Quadros 7, 8 e 9. As tabelas completas estão no Apêndice 3.

QUADRO 7 - ANÁLISE ESTATÍSTICA DESCRITIVA: PARTE I - PERFIL DOS PESQUISADORES

Medidas	Valores	Análise
Desvio Padrão	Entre 0,725 e 2,708	Os valores dentro do intervalo apresentaram coerência nas amostras.
Assimetria	Entre -0,110 e 2,382	Os valores são considerados assimétricos (assimetria negativa).
Curtose	Entre -1,614 e 6,472	Valores indicam distribuição equilibrada e uniforme.

FONTE: Dados da pesquisa (2020).

QUADRO 8 - ANÁLISE ESTATÍSTICA DESCRITIVA: PARTE II - COMPORTAMENTO DOS PESQUISADORES QUANTO AO USO DE DADOS

Medidas	Valores	Análise
Desvio Padrão	Entre 0,687 e 2,454	Os valores dentro do intervalo apresentaram coerência nas amostras.
Assimetria	Entre -2,088 e 0,2855	Os valores são considerados assimétricos (assimetria negativa).
Curtose	Entre -1,250 e 3,221	Valores indicam distribuição equilibrada e uniforme.

FONTE: Dados da pesquisa (2020).

QUADRO 9 - ANÁLISE ESTATÍSTICA DESCRITIVA: PARTE III - COMPORTAMENTO DOS PESQUISADORES QUANTO AO REUSO DE DADOS

Medidas	Valores	Análise
Desvio Padrão	Entre 0,254 e 2,623	Os valores dentro do intervalo apresentaram coerência nas amostras.

Assimetria	Entre -3,489 e 1,476	Os valores são considerados assimétricos (assimetria negativa).
Curtose	Entre -1,991 e 10,457	Uma parte dos valores apresenta distribuição equilibrada e uniforme.

FONTE: Dados da pesquisa (2020).

O coeficiente Alfa de Cronbach e a análise estatística descritiva referente às partes I, II e III indicam que a amostra representa coerência entre si e está dentro de padrão aceitável de distribuição. Os desvios-padrão e variância foram altos, apontando para variabilidade nas respostas dos pesquisadores. A apresentação dos resultados será descrita na seção seguinte.

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Nesta seção são apresentados e discutidos os resultados obtidos durante as fases de coleta, tabulação e análise dos dados à luz do problema de pesquisa: quais são os fatores de impacto percebido pelos pesquisadores sobre o reuso de dados científicos de pesquisa?

Para responder à questão, são apresentados os resultados obtidos na identificação e descrição dos três repositórios de dados exclusivamente brasileiros cadastrados no Re3data.org, em atendimento ao objetivo específico “b”. Na sequência, são apresentados os resultados do obtidos na coleta de dados por meio do questionário que atenderam ao objetivo específico “c”, identificar os pesquisadores que apresentam *datasets* depositados nos repositórios brasileiros, e objetivo específico “d” quanto a categorização dos fatores de impacto percebido pelos pesquisadores quanto ao reuso de dados científicos.














4.1 DESCRIÇÃO DOS REPOSITÓRIOS DE DADOS BRASILEIROS

Para a identificação e descrição dos três repositórios exclusivamente brasileiros cadastrados no R3data.org foi criado um protocolo foi com cinco categorias e 22 critérios levantados com base no referencial teórico sobre repositórios, conforme destaque abaixo:

- Características essenciais aos repositórios de dados (armazenamento, descrição e apresentação dos itens): baseado em Rodrigues et al. (2010) e Paganine e Costa (2016);
- Contemplação do alinhamento do repositório em relação aos critérios da Panton Principles (permissão de uso, licença adequada para dados, sem limitação para uso comercial e demais restrições, e licença do tipo PDDL ou CC0): baseado em Murray-Rust et al. (2010); e
- Atendimento aos requisitos de disponibilidade de informações do Re3data, conforme Figura 5 (informação, acesso, licenças (dados), identificador persistente, certificados e padrões, e política): baseado em Re3data, Pampel et al. (2013) e Vierkant et al. (2014).

A coleta de dados para identificação e descrição dos repositórios foi levantada por meio de consulta no sítio eletrônico dos repositórios, nas políticas de dados desses repositórios e no sítio do Re3data.org. conforme Figura 9.

FIGURA 9 - REQUISITOS DE DISPONIBILIDADE DO RE3DATA.ORG

	ICONE	SIGNIFICADO
Informação		O repositório disponibiliza informações adicionais sobre os seus serviços
Acesso		O repositório oferece acesso aberto aos seus dados
		O repositório oferece acesso restrito aos seus dados
		O repositório oferece acesso fechado aos seus dados
Licenças		Os termos de uso e licenças dos dados são disponibilizados pelo repositório
Identificador Persistente		O repositório usa DOI para tornar seus dados persistentes, únicos e citáveis
		O repositório usa URN para tornar seus dados persistentes, únicos e citáveis
		O repositório usa ARK para tornar seus dados persistentes, únicos e citáveis
		O repositório usa HANDLES para tornar seus dados persistentes, únicos e citáveis
		O repositório usa PURL para tornar seus dados persistentes, únicos e citáveis
		O repositório usa outros esquemas de identificação para tornar seus dados persistentes, únicos e citáveis
Certificados e Padrões		O repositório é certificado ou segue os padrões para repositórios
Política		O repositório possui e disponibiliza um documento de política

FONTE: Sayão e Sales (p. 110, 2016) e Re3ata (s/d, não paginado).

O critério “Licença” dos dados repete-se em *Panton Principles* e nos requisitos de disponibilidade do Re3data.org e por isso foi reunido em um único critério como forma de evitar repetições. O mesmo ocorre para “Identificador persistente” e “Proveniência²¹”.

O primeiro repositório a ser descrito é a BDC/UFPR, seguido do PPBio Data Repository e por fim, IBICT Dataverse Network.

A Base de Dados Científicos (BDC) da Universidade Federal do Paraná foi criada em dezembro de 2017 fruto de uma parceria entre o Centro de Computação

²¹ No contexto desta pesquisa, e no contexto dos Princípios FAIR, a proveniência faz alusão tão somente a indicação ou citação de metadado de ligação que informe a fonte de onde derivam os dados, e não a proveniência de dados em sistemas computacionais (no sentido do conjunto de processos que originam os dados, tais como derivações, parâmetros, artefatos de *hardware* e *software* – *workflow* científico).

Científica e *Software* Livre (C3SL), um laboratório do Departamento de Informática e o Sistema de Bibliotecas (SiBi) da UFPR, repetindo a parceria realizada no desenvolvimento das bibliotecas digitais de teses e dissertações, Biblioteca Digital de Periódicos e Biblioteca Digital de Imagem e Som, constituintes do Repositório Digital Institucional (RDI) da UFPR lançado em 2004. A aplicação do protocolo de análise descritiva no repositório da BDC/UFPR encontrou os seguintes resultados:

QUADRO 10 - PROTOCOLO BDC/UFPR

Critérios de avaliação			BDC/UFPR	
Armazenamento		Software	DSpace	
		Preservação	Backup	
		URL persistente	DOI	
Descrição (metadados)		Administrativos	Sim	
		Estruturais	Sim	
		Técnicos	Não	
		De preservação	Não	
		Tipo de recurso	Dado científico	
		Quantidade de <i>datasets</i>	30	
		Apresentação dos itens		Variedade de recursos
Povoamento	Sim			
Proveniência dos dados	Sim (dc.relation.ispartof)			
Panton Principes				1
		2	Licença adequada para dados	Parcialmente
		3	Sem limitação para uso comercial e demais restrições	Sim
		4	Licença PDDL ²² ou CC0 ²³	Não
Requisitos de disponibilidade		Re3data	Informação	Sim
			Acesso	Sim
			Licenças (dados)	Sim
			Certificados e padrões	Sim
			Política	Sim ²⁴

FONTE: Elaborado pela autora com base nas informações da BDC/UFPR.

A página inicial da BDC/UFPR apresenta descrição do repositório, *link* para acesso à política de gestão de dados que contém informações sobre procedimentos de submissão. A política de gestão de dados indica *link* para uma ferramenta própria para a elaboração do Plano de Gestão de Dados aos pesquisadores. Todas as páginas apresentam informação de contato e suporte via *e-mail* e formulário.

A estrutura da BDC segue o padrão do *software* DSpace e é constituída por uma comunidade que abriga três subcomunidades. Estas subcomunidades dividem-se em sete coleções que abrigam *datasets* individuais ou provenientes de projetos

²² Open Data Commons Public Domain Dedication and Licence (PDDL).

²³ Creative Commons CC Zero (CC0).

²⁴ https://portal.ufpr.br/documentos/BDC/diretrizes_BDC.pdf. Acesso em: 06 mar. 2020.

específicos. Existem coleções para abrigar dados por projetos, dados individuais, dados de grupos de pesquisa e também coleção para dados provenientes dos periódicos editados na Biblioteca Digital de Periódicos (BDP) da universidade, embora esta última não tenha ainda dados depositados no momento desta análise.

As comunidades, subcomunidades e coleções permitem parametrização individualizadas de acordo com a necessidade dos usuários da plataforma, bem como a possibilidade da criação de perfis específicos para atividades como submissão, revisão, edição, publicação e embargo de dados ou conjuntos de dados. Os dados são depositados via autossubmissão mediante cadastro prévio solicitado por *e-mail*. De acordo com as Diretrizes da BDC (2018), a submissão de dados é restrita para a comunidade acadêmica vinculada à universidade e seus egressos, bem como grupos e projetos de pesquisa institucionais; submissões externas devem ser analisadas pelo Comitê Gestor.

As Diretrizes da BDC apresentam um esboço de política de gestão de dados apresentando um pequeno glossário, a descrição da estrutura, do funcionamento, dos tipos e formatos de dados, questões de embargo, dados sigilosos ou sensíveis, instruções para submissão, suporte técnico e indicação de ferramenta própria para elaboração do Plano de Gestão de Dados²⁵ para os pesquisadores. O Plano de Gestão de Dados da BDC/UFPR é baseado nos modelos da DPMTTool e permite edições e versionamentos.

Com relação ao povoamento, a BDC apresenta *datasets* em diversos formatos, bem como relatórios de pesquisas. Os 30 *datasets* levantados na amostra foram confirmados como dados científicos segundo a literatura. Os dados apresentam informações individualizadas sobre licenciamento. Como repositório que se propõe multidisciplinar, a BDC apresenta dados em diversas áreas do conhecimento, com predominância em Ciências Sociais Aplicadas.

Com relação aos dados, observou-se que a página de resultados apresenta informação resumida sobre o recurso com metadados básicos de título, autoria, data de criação, assunto (termos descritores), tipologia do dado, resumo, tamanho dos arquivos e informação sobre licenciamento (FIGURA 10). A maioria dos *datasets* estão documentados (instruções de uso). É possível verificar a descrição completa

²⁵ PGD/UFPR. Disponível em:

https://docs.google.com/forms/d/e/1FAIpQLSfKvTdgJn9bYgX0KfouhNSkM4qKJP9aO1S_yZnkHzUuK0l5GA/viewform. Acesso em: 06 mar 2020.

dos metadados. A BDC utiliza o DOI como identificador persistente para os seus dados.

FIGURA 10 - PÁGINA DE RESUMO DE UM DATASET DA BDC/UFPR

NMR data from the compounds described in the paper "Inhibition of Macrophage Activation and Lymphocyte Function of *Annona vepretorum* Mart. (Annonaceae) Natural Products"

No Thumbnail

This work contain all original NMR data (1D and 2D) from three ent-kaurane diterpenes found in the roots of *Annona vepretorum* (Annonaceae).

DOI: <http://dx.doi.org/10.5380/bdc/13>

Visualizar/Abrir

- NMR data for compound ent-3 β -hydroxykaur-16-ene (3.772Mb)
- NMR data for compound ent-kaur-16-en-19-oic acid (4.321Mb)
- NMR data for compound ent-16 β -hydro-kauran-17-oic acid (4.651Mb)

Este recurso está relacionado com:
<http://dx.doi.org/10.23937/2378-3672/1410021>

Coleções
 Dados [25]

Data
 2018-01-05


Autor
[Santos, Tatiana Barbosa](#)
[Meira, Cássio Santana](#)
[Miranda, Cibele do Carmo](#)
[Menezes, Leonilay Rocha](#)
[Alencar](#)
[Dutra, Lívia Macedo](#)
[Seabra, Líviane do Nascimento](#)
[Barison, Anderson](#)
[Costa, Emmanoel Vinícius](#)
[Guimarães, Elisavira Teixeira](#)
[Seabra, Milena Botelho Pereira](#)

Metadados
[Mostrar registro completo](#)

Assunto
 Annona vepretorum
 NMR
 Diterpenes

Tipo
 Spectral data

Os arquivos de licença a seguir estão associados a este item:
[Creative Commons](#)

 Exceto quando indicado o contrário, a licença deste item é descrito como CC0 1.0 Universal

FONTE: BDC/UFPR (2020).

Quanto ao tipo de licenciamento dos dados, apenas 20 possuem licença do tipo CC0; duas são do tipo *cc-by* (requerem apenas citação do autor original) e oito possuem licença *cc-by-nc-nd* (o tipo mais restritivo que impede a criação de obras derivadas e uso comercial). Esses oito *datasets*, embora disponíveis de forma aberta, não se configuram como dados científicos de pesquisa abertos, pois foram atrelados a um tipo de licença que não permite reuso. Alguns *datasets* são documentados com termos para reuso de dados.

O formato de navegação do repositório segue padrão comum do *software* DSpace com possibilidades de pesquisa básica e avançada, bem como filtros rápidos para recuperação. O repositório apresenta estatísticas de uso contemplando total de visitas, total de visitas por mês, arquivos do *dataset* visitados, visitas por país e cidade.

De forma geral, o repositório de dados da BDC/UFPR atende parcialmente às recomendações para repositórios levantados na literatura. Observou-se que há proporção considerável de dados com licenças restritivas, o que deixa de atender aos propósitos do compartilhamento e reuso de dados.

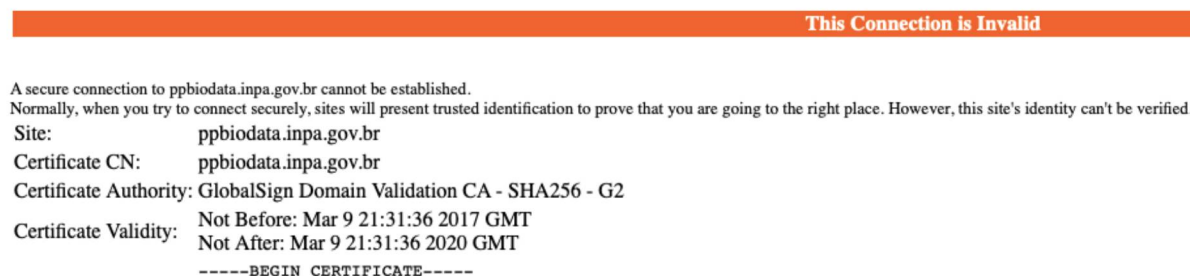
O segundo repositório descrito é o do Programa de Pesquisa em Biodiversidade (PPBio) pertencente ao Instituto Nacional de Ciência e Tecnologia - Centro de Estudos Integrados da Biodiversidade Amazônica (INCT-CENBAM), por sua vez subordinado ao Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTI). O PPBio *Data Repository* desenvolvido em 2004 com a criação do Programa de Pesquisa em Biodiversidade.

O objetivo do PPBio *Data Repository* é “disponibilizar um banco de dados sobre a biodiversidade Amazônica para diversos setores da sociedade” (CENBAM - PPBio, s/d, não paginado). É um repositório de dados bem estruturado e que segue o padrão de metadados *Ecological Metadata Language* (EML) para descrição dos dados. Possui uma política de gestão de dados²⁶ estruturada indicando as responsabilidades do Comitê Gestor de informação, diretrizes sobre gestão, autoria, acesso e uso de dados, bem como os aspectos de propriedade intelectual.

²⁶ Política de dados do PPBio. Disponível em: https://ppbio.inpa.gov.br/sites/default/files/politica_dou.pdf. Acesso em: 09 mar. 2020.

Cabe salientar que no momento da análise descritiva do repositório, o PPBio *Data Repository* estava indisponível devido a certificado inválido do sítio eletrônico, conforme Figura 11.

FIGURA 11 - INDISPONIBILIDADE DO PPBIO



FONTE: Sítio eletrônico do PPBio acessado em 09 mar. 2020.

Assim sendo, considerando que toda a interface e ambiente do repositório PPBio é o mesmo da plataforma *DataONE*, local para onde o repositório PPBio é replicado, a análise foi realizada sem prejuízos na coleção “PPBio” no *DataONE* e os resultados obtidos foram explicitados no quadro abaixo.

QUADRO 11 - PROTOCOLO PPBIO DATA REPOSITORY

Critérios de avaliação		PPBio <i>Data Repository</i>
Armazenamento	<i>Software</i>	<i>MetaCat</i>
	Preservação	Replicação e cópia para o <i>DataONE</i>
	URL persistente	URN
Descrição (metadados)	Administrativos	Sim
	Estruturais	Sim
	Técnicos	Sim
	De preservação	Não
	Tipo de recurso	Dado científico
Apresentação dos itens	Quantidade de <i>datasets</i>	535
	Variedade de recursos	Sim
	Povoamento	Sim
	Proveniência dos dados	Sim
Panton Principes	1 Permissão de uso	<i>cc by-nc-nd</i>
	2 Licença adequada para dados	Não
	3 Sem limitação para uso comercial e demais restrições	Não
	4 Licença PDDL ou CC0	Não
Requisitos de disponibilidade	Re3data	Informação
		Acesso
		Licenças (dados)
		Certificados e padrões
		Política

FONTE: Elaborado pela autora com base nas informações do PPBio (2020).

O PPBio utiliza-se da estrutura do diretório *DataONE* suportada pelo *software* livre *MetaCat*²⁷ para gestão dos dados científicos. A interface de busca e recuperação segue os padrões do *DataONE* que apresentam estrutura simplificada para uso. Todas as páginas possuem campo para pesquisa básica e avançada e um menu lateral com diversas opções de filtros rápidos, incluindo filtro de pesquisa por taxonomia. A interface permite customização da exibição dos resultados (mais recente, por identificador, título e autor – na ordem crescente e/ou de A a Z). O repositório permite o versionamento dos dados e projetos e apresenta contador de citações, *downloads*, visualizações e possui mecanismo para copiar a citação pronta (FIGURA 12).

Quanto ao povoamento, o PPBio não apresenta separação por coleções. Os dados são apresentados na ordem de publicação. Dos 535 dados levantados inicialmente, todos eles foram confirmados como dados científicos de pesquisa contendo metadados completos e padronizados, embora alguns dados estivessem sob embargo. De acordo com sua Política de Dados, todos os *datasets* possuem informações padronizadas devido a utilização obrigatória do *software Morpho*²⁸ de apoio à construção de metadados de acordo com o padrão EML. O PPBio disponibiliza em sua página tutorial no formato texto e vídeo para uso do *Morpho*.

FIGURA 12 - RESUMO PPBIO DATA REPOSITORY



FONTE: PPBio Data Repository (2020).

²⁷ Knowledge Network for Biocomplexity. Disponível em: <https://knb.ecoinformatics.org/knb/docs/>. Acesso em: 09 mar. 2020.

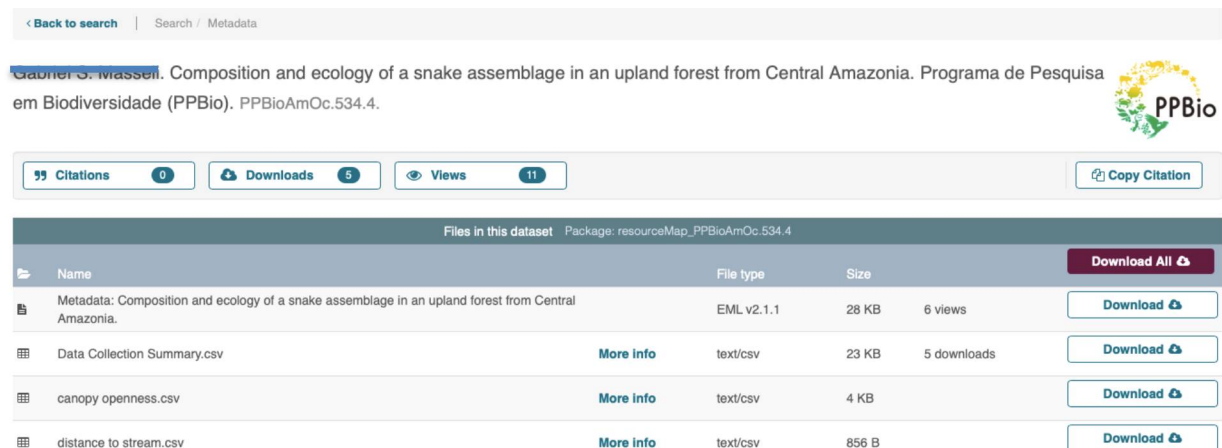
²⁸ Tutorial do *Morpho*. Disponível em: <https://knb.ecoinformatics.org/software/dist/MorphoUserGuide.pdf>. Acesso em: 09 mar. 2020.

Em relação a submissão, os *datasets* são preparados pelos pesquisadores com auxílio do *Morpho* e tabelas pré-definidas pela gestão do PPBio e enviados para a equipe do Comitê Gestor de Informação que fará a submissão dos *datasets* no repositório. O prazo para divulgação pública dos dados é até um ano após a coleta pelo pesquisador e de sete dias para disponibilização pública. No tocante a direitos autorais e licenças de uso, a política de dados indica que a propriedade intelectual dos dados coletados no âmbito das pesquisas no CENBAM-PPBio pertence aos pesquisadores-autores e é de “responsabilidade do Ministério da Ciência e Tecnologia” (POLÍTICA DE DADOS PPBio, 2009, p. 9). O item 4.2 da política indica que

os dados coletados, gerados, e disponibilizados no âmbito do PPBio são de utilização prioritária para fins educacionais, culturais, científicos, de divulgação e de gestão pública. O acesso e uso com intenção comercial ou de forma que possa resultar na geração de produtos ou processos passíveis de exploração econômica, deverá ocorrer mediante a celebração de contrato entre as partes interessadas, observada a legislação pertinente e as disposições dessa política (POLÍTICA DE DADOS PPBio, 2009, p. 9).

O padrão de licença para os dados é o mais restritivo possível, se aproximando de um padrão *Creative Commons* cc-by-nc-nd, visto que não permite uso, modificações, compartilhamento e reuso sem permissão expressa pelos autores e pelo PPBio. Na Figura 13 observa-se o exemplo de um *dataset* com destaque na declaração expressa de licenciamento:

FIGURA 13 - EXEMPLO DE DATASET DO PPBio DATA REPOSITORY



< Back to search | Search / Metadata

Gabriel S. Massen. Composition and ecology of a snake assemblage in an upland forest from Central Amazonia. Programa de Pesquisa em Biodiversidade (PPBio). PPBioAmOc.534.4.

PPBio

Citations 0 Downloads 5 Views 11 Copy Citation

Files in this dataset Package: resourceMap_PPBioAmOc.534.4

Name	File type	Size	Download All
Metadata: Composition and ecology of a snake assemblage in an upland forest from Central Amazonia.	EML v2.1.1	28 KB 6 views	Download
Data Collection Summary.csv	text/csv	23 KB 5 downloads	Download
canopy openness.csv	text/csv	4 KB	Download
distance to stream.csv	text/csv	856 B	Download

(...)

Data Set Usage Rights

Intellectual Rights This dataset is released to the public and may be freely downloaded. Please keep the designated Contact person informed of any plans to use the dataset. Consultation or collaboration with the original investigators is strongly encouraged. Publications and data products that make use of the dataset must include proper acknowledgement. For more information on PPBio Network data access and use policies, please see: http://ppbio.inpa.gov.br/Port/docsinternos/politica_dou.pdf.

Access Control

Auth System

Order

Allow

Permission	Principal
[read]	public
[all]	uid=PPBioAmOc,o=PPBio-CENBAM,dc=ecoinformatics,dc=org

FONTE: PPBio (2020).²⁹

Nesse sentido, observou-se que embora os dados do PPBio estejam em acesso aberto, eles são restritos quanto a modificações, compartilhamento e reuso, o que demonstra ainda não estar completamente alinhado às práticas da ciência aberta e à definição de um dado científico aberto conforme a literatura.

O terceiro e último repositório analisado foi o IBICT *Dataverse Network*, do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), unidade de pesquisa pertencente ao MCTI. O IBICT *Dataverse* foi registrado no Re3data.org em outubro de 2016. O repositório do IBICT foi criado para “garantir acesso continuado em longo prazo dos conteúdos armazenados digitalmente” (ARAÚJO; MÁRDERO ARELLANO; FERRER, 2018, p. 6) para os dados de pesquisa da Rede Brasileira de Serviços de Preservação Digital Cariniana.

A Rede Cariniana é uma rede de preservação digital de informação científica a longo prazo que armazena publicações científicas ligadas às plataformas *Open Journal System/Sistema Eletrônico de Editoração de Revistas (OJS/SEER)*, Portal do Livro Aberto e Biblioteca Digital de Teses e Dissertações (BDTD) e, posteriormente, das teses e dissertações armazenadas no *software DSpace* das instituições parceiras ligadas à Cariniana.

O IBICT *Dataverse Network* utiliza o *software Dataverse* desenvolvido pela *Harvard University*. É um *software* livre para publicação, citação, análise e preservação de dados científicos de pesquisa. A organização do *Dataverse* compreende seis níveis onde é possível o armazenamento de *datasets* por área do conhecimento, instituições, publicações, projetos, pesquisas institucionais e pesquisas individuais. Cada nível permite customização individual de parâmetros e

²⁹ <https://search.dataone.org/view/PPBioAmOc.534.4>. Acesso em: 07 mar. 2020.

permissões de acesso. No caso de projetos, por exemplo, há uma coleção específica para revistas científicas ligadas à Rede Cariniana onde é possível aos editores submeterem os dados científicos provenientes dos artigos publicados ou *data papers*.

A aplicação do protocolo de análise apresentou os dados observados no quadro abaixo:

QUADRO 12 - PROTOCOLO IBICT *DATaverse NETWORK*

Critérios de avaliação		IBICT <i>Dataverse Network</i>
Armazenamento	<i>Software</i>	<i>Dataverse</i>
	Preservação	LOCKSS
	URL persistente	DOI
Descrição (metadados)	Administrativos	Sim
	Estruturais	Sim
	Técnicos	Sim
	De preservação	Não
	Tipo de recurso	Publicações científicas, relatórios, dado científico
	Quantidade de <i>datasets</i>	153
	Proveniência	Não
Apresentação dos itens	Variedade de recursos	Sim
	Povoamento	Sim
	Proveniência dos dados	Sim
Panton Principes	1	Permissão de uso
	2	Licença adequada para dados
	3	Sem limitação para uso comercial e demais restrições
	4	Licença PDDL ou CC0
Requisitos de disponibilidade	Re3data	CC0, <i>cc-by</i> e restrito
		Parcialmente
		Sim
		Sim, parcialmente
		Sim
		Sim
		Parcialmente
		Sim
		Não

FONTE: Elaborado pela autora com base nas informações do IBICT *Dataverse Network* (2020).

O IBICT *Dataverse* possui menu com acesso à documentação do *software*, ajuda automática e guia de uso da plataforma *Dataverse*. O suporte é realizado por meio de formulário eletrônico. Apresenta sistema de busca simples e busca avançada com filtros rápidos para recuperação. Os seis níveis da estrutura do IBICT *Dataverse* permitem a criação de “dataversers” dentro do repositório para abrigar projetos e/ou coleções específicas. É dentro desses “dataversers” que os *datasets* são submetidos na plataforma e ficam atrelados aos projetos (FIGURA 14).

FIGURA 14 - PÁGINA RESUMO IBICT DATAVERSE NETWORK



FONTE: Portal da Rede Cariniana (2019).

Existem 14 “dataverses” com um total de 153 *datasets*. Não foi possível verificar individualmente cada *dataset* devido a indisponibilidade do repositório na web desde meados de fevereiro 2020 e até o fechamento desta pesquisa não foi obtido retorno do *e-mail* encaminhado ao administrador do serviço.

Considerando os dados que foram observados no momento de definição da amostra, constatou-se que nem todos os conjuntos de dados eram de fato dados científicos de pesquisa. Dentro de “dataverses” de projetos específicos com o do Jardim Botânico de Brasília, todos os conjuntos de dados eram, na realidade, relatórios da administração do Jardim Botânico sem dados científicos apensados. Nesse sentido, embora a descrição do repositório no Re3data.org, no sítio da Rede Cariniana e no “Guia para usuários do repositório *Dataverse* do IBICT” (ARAÚJO; MÁRDERO ARELLANO; FERRER, 2018) descreva-o como repositório de dados, não foi possível comprovar seu total enquadramento conforme o referencial teórico devido a indisponibilidade do recurso na web.

Com relação a submissão de dados, é requerido um cadastro prévio no repositório e apresentar vinculação de projeto de pesquisa junto ao IBICT. Após análise, a submissão estará disponível somente para pesquisadores com projetos

que façam parte da rede de pesquisadores do IBICT, do grupo de pesquisa Dríade³⁰, pesquisadores de instituições parceiras da Rede Cariniana ou periódicos registrados na plataforma OJS/SEER vinculados à Rede.

Com relação às suas funcionalidades, o repositório apresenta ferramentas adequadas de gestão para administradores e curadores de dados, permitindo customização de temas, cores, adição de imagens; recursos para citação (XML para o gerenciamento *EndNote*, formatos RIS e BibTeX). Há possibilidades de permissões de diferentes níveis de acesso para colaboradores e usuários.

Quanto aos dados, o pesquisador depositante pode inserir outros metadados além daqueles já oferecidos no formulário de preenchimento padrão; pode ainda editar, definir o status do *dataset* (se restrito ou irrestrito). Outra função relevante é a possibilidade de criar novas versões dos dados mantendo as versões anteriores, o que enriquece o material à medida que novas análises são realizadas e novos descobrimentos são realizados.

4.2 PERFIL DOS PESQUISADORES E CATEGORIZAÇÃO DOS FATORES DE IMPACTO

Com a finalidade de identificar o perfil dos pesquisadores e categorizar os fatores de impacto percebidos pelos pesquisadores quanto ao reuso de dados, foi aplicado um questionário com 25 perguntas fechadas.

O questionário recebeu 73 respostas. Às respostas foram aplicadas análises estatísticas para atestar a confiabilidade e a conformidade da distribuição dos dados. O questionário foi dividido em três partes com objetivos distintos de levantar e identificar:

1. o **perfil** institucional dos pesquisadores, a fim de verificar a origem das pesquisas, tipo de financiamento, área do conhecimento no qual atuam e qual(ais) é (são) o(os) tipo(os) de dados que trabalham;
2. o comportamento dos pesquisadores quanto ao **uso de dados científicos**, objetivando identificar o nível de profundidade no trabalho destes pesquisadores com dados científicos; e

³⁰ Grupo de pesquisa em preservação digital ligado ao IBICT. Disponível em: <http://dgp.cnpq.br/dgp/espelhogrupo/3997875180380796>. Acesso em: 06 mar. 2020.

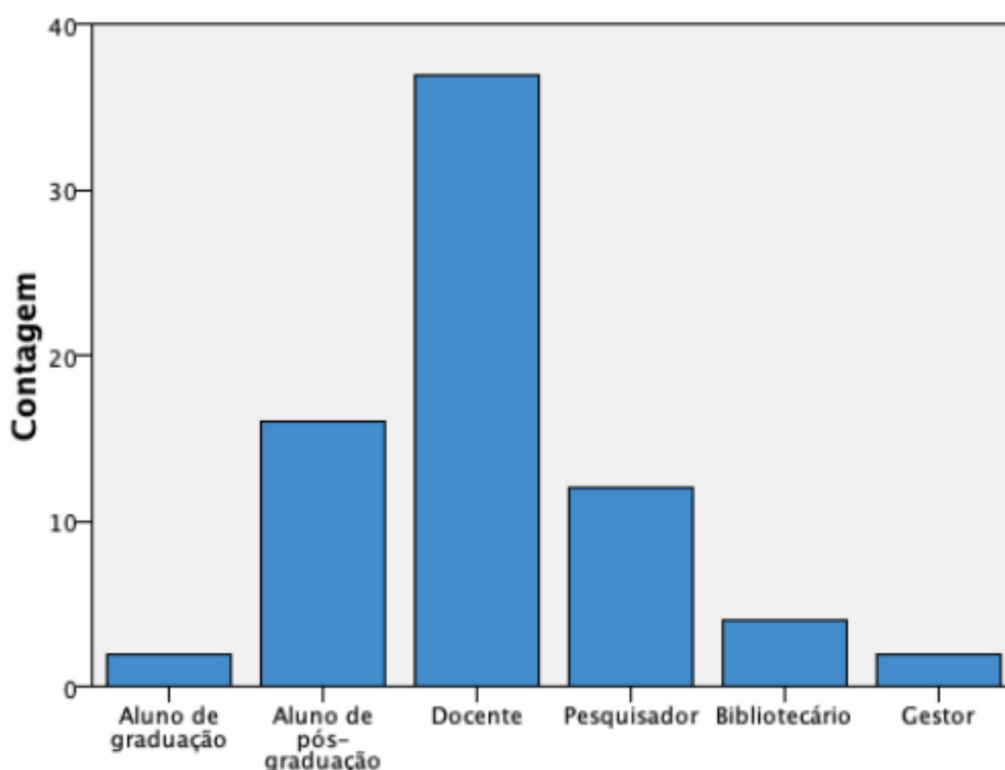
3. o comportamento dos pesquisadores quanto ao **reuso de dados** científicos, com propósito de identificar como o reuso de dados impacta na atividade acadêmica/científica desses pesquisadores.

Os resultados foram agrupados e analisados de acordo com a divisão acima explicitada e foram descritos nas subseções abaixo.

4.2.1 Perfil dos pesquisadores

Com o objetivo de caracterizar a amostra, foi levantado o perfil do pesquisador, situando-o conforme seu grau formal de instrução educacional. Verifica-se que a principal ocupação é a docência com 50,7% dos resultados. Em seguida, estão os alunos de pós-graduação com 21,9%, seguidos pelos pesquisadores com 16,4%, conforme Gráfico 2.

GRÁFICO 2 - PRINCIPAL OCUPAÇÃO DOS RESPONDENTES

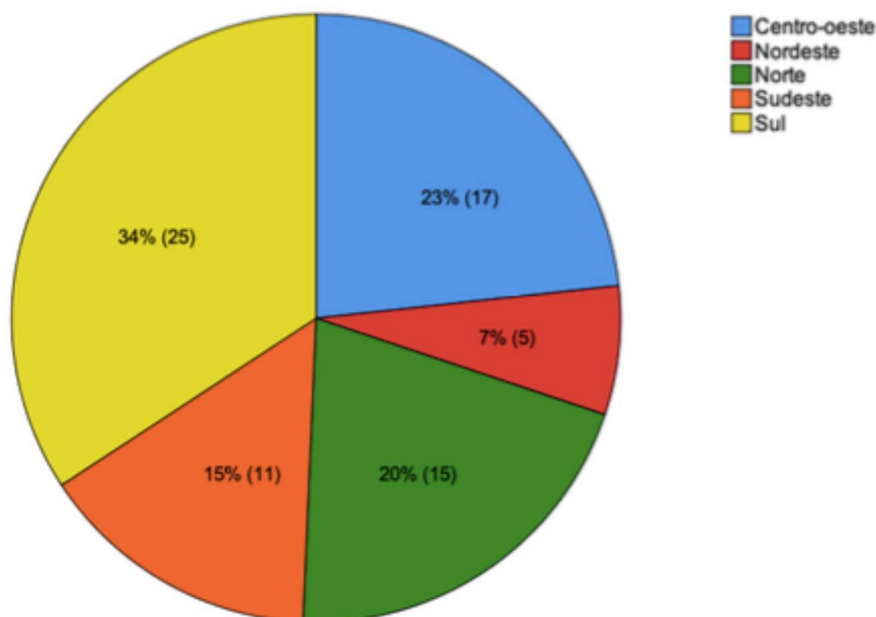


FONTE: Dados da pesquisa (2020).

Com relação à distribuição dos pesquisadores por região, de acordo com o Gráfico 3, observa-se a predominância na região Sul do Brasil, com 34% dos

respondentes. Em seguida aparece a região Centro-oeste com 23%, e a região Norte com 20,5%, respectivamente.

GRÁFICO 3 - PRINCIPAL LOCAL DE TRABALHO DOS PESQUISADORES



FONTE: Dados da pesquisa (2020).

Ao comparar o resultado obtido pela pesquisa com o último censo realizado pelo CNPq em 2016³¹ sobre a distribuição de pesquisadores brasileiros por região, nota-se que o perfil dos pesquisadores que disponibilizam e reusam dados científicos desloca-se da distribuição dos pesquisadores levantados pelo censo, pois o relatório do CNPQ apresenta que a maior concentração de pesquisadores está na região Sudeste (74,4%), seguida da região Sul (67,4%) e Centro-Oeste (67%).

Em relação à área do conhecimento de atuação dos pesquisadores, observou-se que a maioria dos respondentes se concentra na área de Ciências Biológicas (31,5%). As áreas de Ciências Sociais Aplicadas e Ciências Ambientais aparecem na segunda e terceira colocação, respectivamente conforme Tabela 5. Apenas 1 pesquisador atua na área de Bioética. Importante ressaltar que as grandes

³¹ Distribuição de pesquisadores segundo a região geográfica [Brasil]. Fonte: CNPq (s/d). Disponível em: <http://lattes.cnpq.br/web/dgp/por-regiao1>. Acesso em: 22 fev. 2020.

áreas do conhecimento selecionadas como opções de resposta para o questionário foram extraídas conforme a tabela de áreas do conhecimento da CAPES³².

TABELA 4 - ÁREA DE ATUAÇÃO

Área do Conhecimento	Frequência	%
Ciências Agrárias	5	6,8%
Ciências Biológicas	23	31,5%
Ciências Humanas	4	5,5%
Ciências da Saúde	3	4,1%
Ciências Sociais Aplicadas	14	19,2%
Engenharias	7	9,6%
Linguística, Letras e Artes	3	4,1%
Bioética	1	1,4%
Ciências Ambientais	13	17,8%
Total	73	100%

FONTE: Dados da pesquisa (2020).

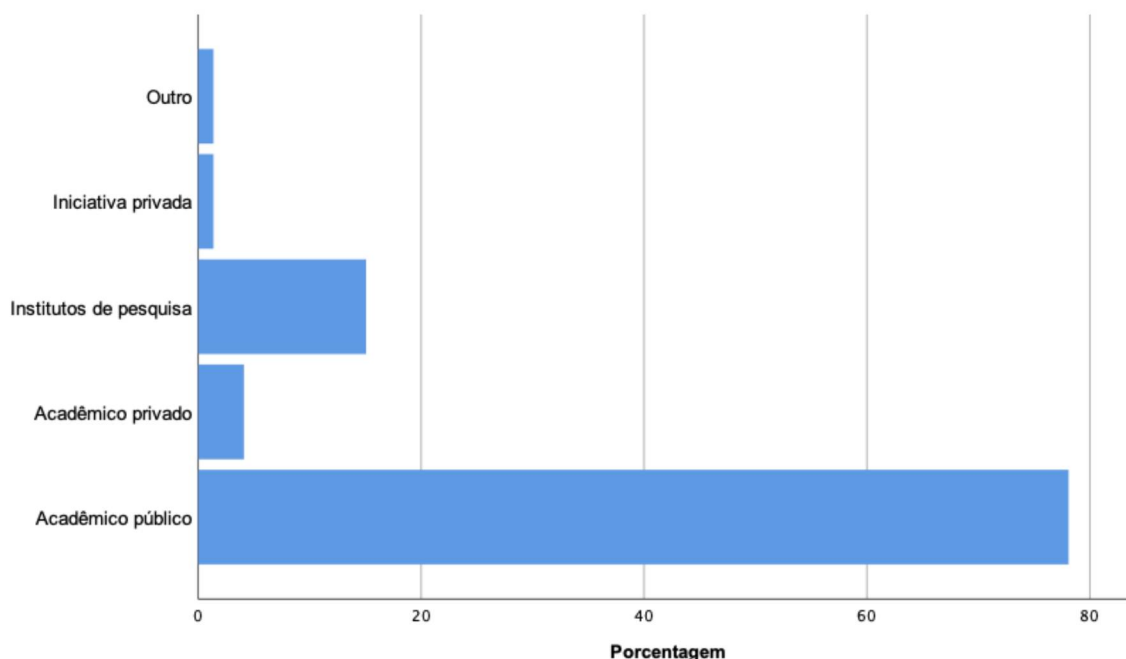
No que tange a área do conhecimento, observou-se o destaque para atuação dos pesquisadores nas disciplinas que compõem as chamadas ciências da vida (ciências biológicas, ambientais, agrárias e saúde), com 61,6% dos resultados.

Os pesquisadores responderam que sua principal abordagem metodológica reside no tratamento misto aos dados (pesquisas qualitativas e quantitativas) com 58,9%. Enquanto as abordagens qualitativas e quantitativas ficaram com 23,3% e 17,8%, respectivamente.

Com relação à vinculação profissional, os pesquisadores são em maioria ligados a instituições acadêmicas públicas, com 78,1% das respostas. Os profissionais vinculados a exclusivamente a institutos de pesquisa representam 15% do total.

³² Tabela de áreas do conhecimento. Fonte: Fundação CAPES (2018). Disponível em: <https://www.capes.gov.br/avaliacao/instrumentos-de-apoio/tabela-de-areas-do-conhecimento-avaliacao>. Acesso em: 10 de out. 2019.

GRÁFICO 4 - SETORES DE VINCULAÇÃO PROFISSIONAL DOS PESQUISADORES



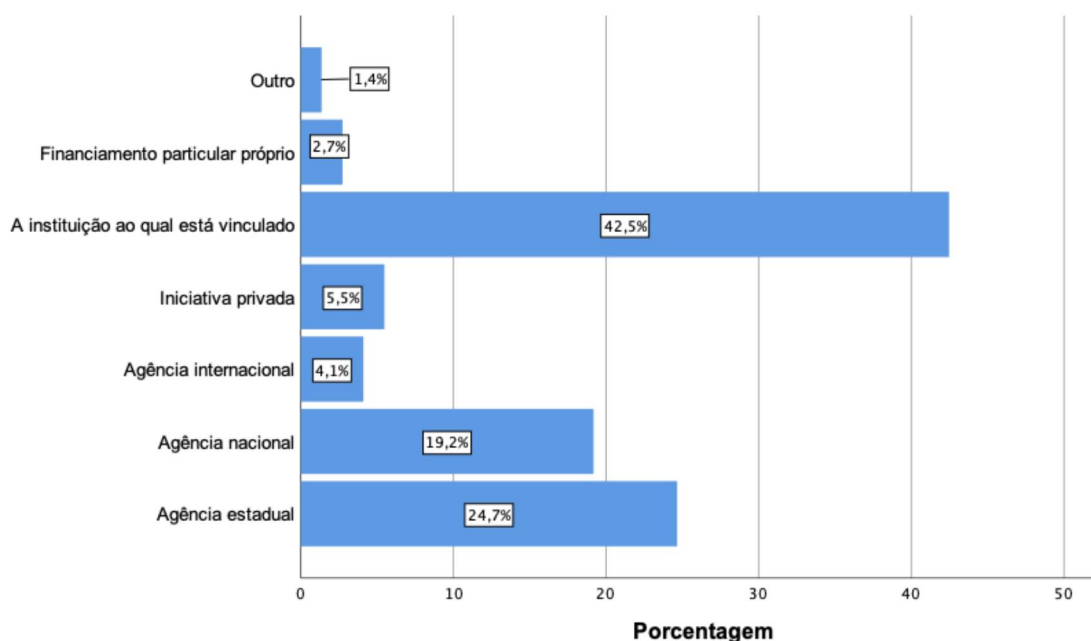
FONTE: Dados da pesquisa (2020).

Nesta questão foi incluída o tipo de resposta “Outros. Especifique” a fim de que o pesquisador pudesse descrever seu vínculo caso sua categoria não estivesse representada nas opções de respostas. Um dos pesquisadores acrescentou que é “Professor de biologia da Rede Estadual de SC”. Este caso, especificamente, não pode ser considerado dentro da opção “Acadêmico público” porque o termo “academia” faz referência ao ambiente do ensino superior, conforme verbete do dicionário Michaelis³³.

Com relação ao tipo de financiamento da pesquisa, os respondentes indicaram que os recursos estão majoritariamente vinculados à instituição profissional em que atuam com 42,5%. Em segundo lugar, aparecem as agências estaduais com 24,7% e agências nacionais com 19,2% das respostas. As demais respostas seguem no Gráfico 5.

³³ FONTE: Michaelis Moderno Dicionário da Língua Portuguesa. Disponível em: <http://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=academia>. Acesso em: 22 de fev. 2020.

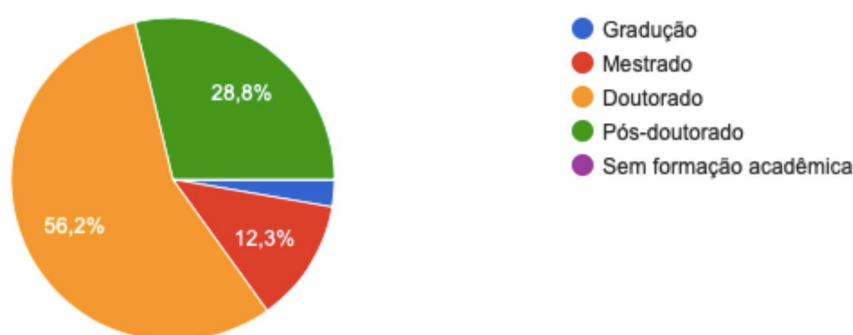
GRÁFICO 5 - PRINCIPAIS FINANCIADORES DE PESQUISA



FONTE: Dados da pesquisa (2020).

Quanto ao nível de escolaridade dos pesquisadores, o Gráfico 6 descreve a formação dos respondentes. A formação em doutor é majoritária entre os respondentes com 56,2%. Em seguida, aparecem a formação pós-doutorado com 28,8%, mestre com 12,3% e graduado com 2,7%.

GRÁFICO 6 - NÍVEL DE ESCOLARIDADE DOS PESQUISADORES



FONTE: Dados da pesquisa (2020).

Embora a questão acima esteja diretamente relacionada com a questão nº1 do instrumento de coleta de dados, essa pergunta foi necessária para identificar de modo geral qual é a formação do pesquisador brasileiro depositário de conjuntos de dados em repositórios nacionais. A maioria dos *datasets* depositados possui autoria

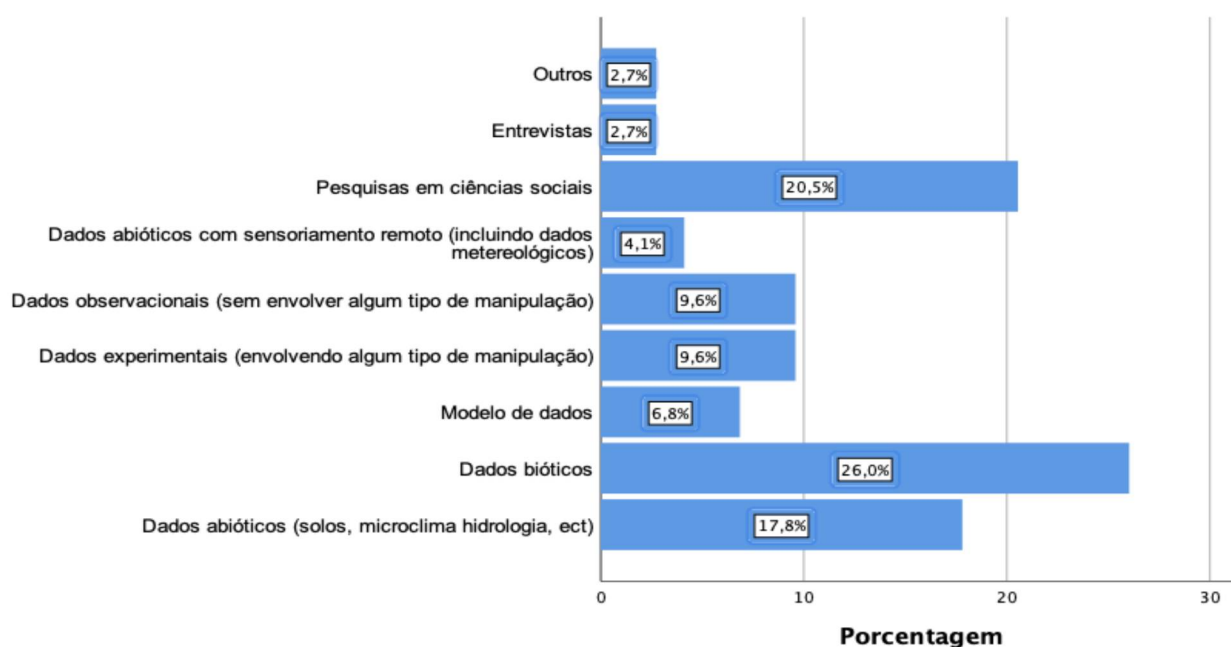
múltipla e supõem-se que as equipes ou grupos de pesquisa possuem docentes e pesquisadores em diversos níveis. Como o questionário foi encaminhado para todos os pesquisadores identificados com nome e *e-mail*, é provável que outros membros além do líder do grupo tenham respondido ao questionário e por isso, a variabilidade entre as respostas obtidas.

4.2.2 Comportamento sobre dados científicos de pesquisa

A Parte II do questionário foi composta por seis questões cujo objetivo foi identificar o nível de profundidade dos pesquisadores quanto ao uso de dados científicos de pesquisa, servindo como comparativo com relação ao comportamento de reuso que será apresentado na próxima subseção.

A questão de nº 8 perguntou qual termo melhor descrevia o tipo de dado utilizado pelos pesquisadores. As respostas indicam os dados bióticos como o tipo de dado mais representativo, com 26%. O segundo tipo de dado manejado pelos pesquisadores são os dados de pesquisa em ciências sociais com 20,5% e os dados abióticos (solos, microclima, hidrologia, etc) com 17,8%. Os dados experimentais (envolvendo algum tipo de manipulação) e observacionais (sem envolvimento de qualquer tipo de manipulação) registraram quase 10% das respostas, respectivamente, conforme gráfico abaixo.

GRÁFICO 7 - TIPO DE DADO UTILIZADO PELOS PESQUISADORES



FONTE: Dados da pesquisa (2020).

Para fins de esclarecimento, segue abaixo descrição sobre os tipos de dados:

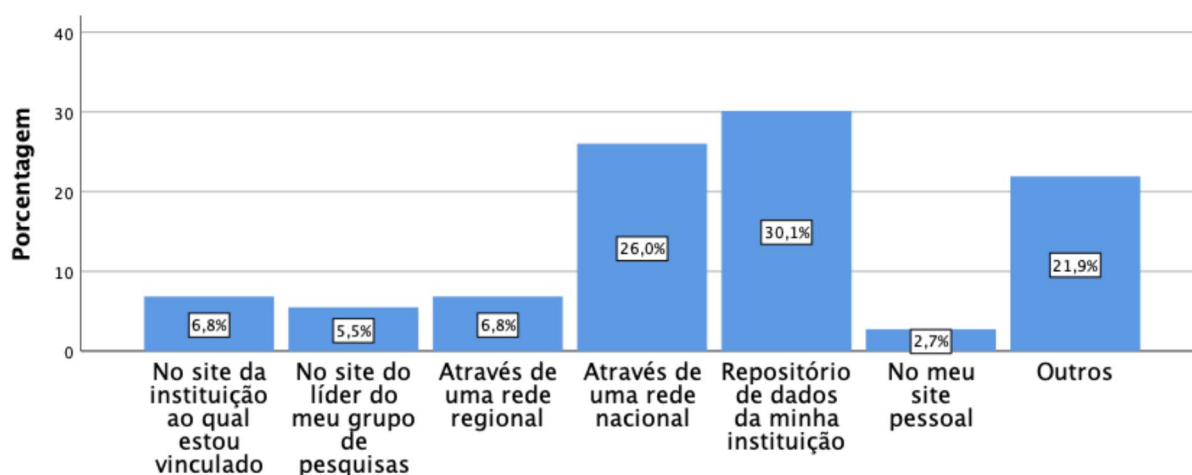
- **dados bióticos** são aqueles que resultam da interação entre os seres vivos de uma determinada região, formando a comunidade biológica que influencia o ecossistema do qual fazem parte, baseados a partir das relações ecológicas e das cadeias alimentares (DIANA, s/d);
- **dados abióticos** são aqueles resultantes de fatores abióticos que influenciam os seres vivos de um ecossistema, podendo prover de aspectos físicos, químicos, físico-químico influenciando diretamente o meio ambiente, tais como: luz, temperatura, vento, água, radiação, pressão, composição do solo etc. (AZEVEDO et al., 2015);
- **dados observacionais** estão ligados a observações diretas associadas a medições, lugar e tempo específicos e podem ser qualificados como registros históricos – dados que não poderão ser coletados por uma segunda vez devido a suas características de coleta em momento específico (SAYÃO; SALES, 2014);
- **dados experimentais** caracterizam-se pelo controle de todas as condições experimentais e suas variáveis durante a realização do experimento; são dados que podem ser testados sem limites desde que preservados os insumos e condições específicas do evento e si (BORGMAN, 2012);
- **modelos de dados** são os dados computacionais resultados da execução de modelos e/ou simulações; precisam da emulação perfeitas dos seus ambientes para serem reproduzidos; e
- **pesquisas em ciências sociais** envolvem abordagens qualitativas e quantitativas a fim de analisar uma gama de diferentes eventos sociais (DEMO, 2008). Os tipos de dados produzidos podem ser dados experimentais, entrevistas, observações, entre outros.

Essa questão incluiu opção de resposta do tipo “Outros. Especifique” para o caso do tipo de dado que o pesquisador manipula não estar arrolado nas opções de resposta. Dessa forma, foram recebidas outras duas respostas por pesquisadores distintos: “tabelas e gráficos (série temporal)” (respondente nº 6) e “dados abertos” (respondente nº 10). Tabelas e gráficos não se caracterizam como um tipo de dado científico de pesquisa, mas sim enquanto formato por meio do qual os dados são dispostos e/ou visualizados. Por “dados abertos” entende-se a forma de

disponibilização de um dado ao público e não a tipologia em si; assim, nenhuma das duas respostas pode ser considerada um tipo de dado científico.

Em relação a disponibilização de dados científicos, foi perguntado aos pesquisadores onde estes estão disponíveis. A maioria respondeu que seus dados estão armazenados em “Repositório de dados de sua instituição” (30,1% das respostas). O segundo local escolhido para depósito dos dados são as “redes nacionais” [de repositório de dados], com 26%, conforme Gráfico 8:

GRÁFICO 8 - LOCAL ONDE OS PESQUISADORES DISPONIBILIZAM SEUS DADOS



FONTE: Dados da pesquisa (2020).

Na opção de resposta do tipo “Outros. Especifique”, surgiram as seguintes respostas além das opções pré-estabelecidas:

- “*Github*”;
- “Geralmente no material suplementar do/dos artigo/s em que os dados são publicados”;
- “Os dados estão em um repositório que também possuem publicações”;
- “Meus dados não estão disponíveis”;
- “Nuvem (compartilhamento)”;
- “*Lattes*”;
- “*Journal*”;
- “*Google Drive*”;
- “Repositório de dados internacional (*DataONE*)”;
- “No site das revistas acadêmicas”; e
- “No site pessoal”.

Ao categorizar as respostas acima, observa-se que três indicam o armazenamento dos dados junto das publicações, ou seja, no sítio eletrônico das revistas acadêmicas junto aos artigos. Outras três respostas concentram o armazenamento em repositórios gerais como *Github*, *Google Drive* e na “nuvem”. Cabe aqui observar que esses três repositórios não possuem caráter acadêmico e sua abertura estão condicionados tão somente ao desejo pessoal dos pesquisadores em torná-los abertos; não estão condicionados a uma política de abertura e disponibilização de dados. Esses repositórios não apresentam identificadores persistentes para suas coleções, podendo variar o endereço eletrônico utilizado para sua recuperação.

Duas respostas indicam a disponibilização de dados em repositórios formais, sendo um deles repositório institucional e/ou temático e o outro, o *DataONE*, uma plataforma de excelência no armazenamento, preservação e divulgação de dados observacionais da Terra mantido pela Fundação Nacional de Ciência dos Estados Unidos. Outra resposta apontou a disponibilização de dados via Plataforma Lattes. Todavia, a Plataforma não apresenta campo específico para submissão de dados científicos. Um dos respondentes afirmou a totalidade dos seus dados não está disponível.

A questão de nº 10 perguntou aos pesquisadores a razão pela qual todos os seus dados não estariam disponíveis abertamente. Nesse sentido, os resultados com maior frequência apontam que os pesquisadores não tiveram tempo suficiente para torná-los disponíveis (8,2%) e a disponibilidade dos dados não era uma exigência do financiador da pesquisa com 5,5%. A ausência de recursos financeiros foi a terceira resposta mais frequente, com 4,1% dos resultados. A Tabela 5 indica o detalhamento das respostas.

TABELA 5 - INDISPONIBILIDADE NA PUBLICAÇÃO DE TODOS OS DADOS

Motivo para indisponibilização	Frequência	%
Ausência de recursos financeiros	3	4,1%
Ausência de padrões para disponibilização	1	1,4%
Desconhecimento de padrões para disponibilização	1	1,4%
Não tive tempo suficiente para disponibilizá-los	6	8,2%
Eles não deveriam estar disponíveis	1	1,4%
O financiador da pesquisa não exige disponibilização de dados	4	5,5%

Não tenho o direito de torná-los públicos	2	2,7%
Todos os meus dados estão disponíveis	55	75,3%
Total	73	100%

FONTE: Dados da pesquisa (2020).

Após verificar situações de disponibilização de dados, as questões seguintes versaram sobre a satisfação do pesquisador com a gestão dos seus dados científicos. Nesse sentido, a questão nº 11 abordou o grau de satisfação dos pesquisadores sobre coleta e uso de dados científicos de pesquisa (TABELA 6).

TABELA 6 - COMO OS PESQUISADORES COLETAM E USAM DADOS CIENTÍFICOS DE PESQUISA (QUESTÃO 11)

Constructos	Mediana	Moda	Soma	Desvio Padrão	Variância
Estou satisfeito com o meu processo de coleta de dados	4,00	4	277	0,763	0,582
Estou satisfeito com o processo para coletar meus dados	3,00	3	253	0,835	0,697
Estou satisfeito com o processo para armazenar meus dados durante a vida do projeto (curto prazo)	4,00	4	266	0,918	0,844
Estou satisfeito com o processo para armazenar meus dados para além da vida do projeto (longo prazo)	4,00	4	265	1,112	1,236
Estou satisfeito com o processo para análise dos meus dados	4,00	4	280	0,687	0,473
Eu compartilho meus dados com os outros	4,00	4	262	1,078	1,1162
Outros podem facilmente acessar meus dados	3,00	3	235	1,044	1,090
Estou satisfeito com as ferramentas para preparar meus dados	4,00	4	250	0,881	0,775
Estou satisfeito com as ferramentas para preparar a documentação dos meus dados	3,00	3	242	0,911	0,830

FONTE: Dados da pesquisa (2020).

De modo geral, quase 70% dos pesquisadores demonstraram satisfação com a forma como coletam, usam e preparam os dados e a documentação desses dados científicos de pesquisa (QUADRO 13). Todavia, mostraram-se neutros quanto a coleta, acesso e preparação de seus próprios dados. Essa neutralidade chama atenção para a provável ausência de algum processo e/ou atividade que garanta um grau a mais de segurança nos processos de preparação, coleta e acesso aos dados dos próprios pesquisadores.

QUADRO 13 - DISTRIBUIÇÃO DE FREQUÊNCIA - COMO OS PESQUISADORES COLETAM E USAM DADOS

Questão 11	Discordo Fortemente	Discordo	Não discordo, nem concordo	Concordo	Concordo Fortemente
Estou satisfeito com o meu processo de coleta de dados	4(5,5%)	5(6,8%)	33(45,2%)	26(35,6%)	5(6,8%)
Estou satisfeito com o processo para coletar meus dados	3(4,1%)	5(6,8%)	28(38,4%)	32(43,8%)	5(6,8%)
Estou satisfeito com o processo para armazenar meus dados durante a vida do projeto (curto prazo)	6(8,2%)	7(33%)	33(45,2%)	19(26%)	8(11%)
Estou satisfeito com o processo para armazenar meus dados para além da vida do projeto (longo prazo)	5(6,8%)	4(5,5%)	21(28,8%)	29(39,7%)	14(19,2%)
Estou satisfeito com o processo para análise dos meus dados	0(0%)	2(2,7%)	18(24,7%)	43(58,9%)	10(13,7%)
Eu compartilho meus dados com os outros	4(5,5%)	9(12,3%)	12(16,4%)	33(45,2%)	15(20,5%)
Outros podem facilmente acessar meus dados	3(4,1%)	4(5,5%)	18(24,7%)	39(53,4%)	9(4,1%)
Estou satisfeito com as ferramentas para preparar meus dados	2(2,7%)	3(4,1%)	34(46,6%)	27(37%)	7(9,6%)
Estou satisfeito com as ferramentas para preparar a documentação dos meus dados	0(0%)	4(5,5%)	18(24,7%)	40(54,8%)	11(15,1%)

FONTE: Dados da pesquisa (2020).

A questão 12 perguntou sobre o envolvimento que a instituição e/ou grupo de pesquisa tem com os dados produzidos pelos pesquisadores. As respostas indicaram que suas instituições e/ou grupos de pesquisa possuem envolvimento neutro quanto a produção de dados (TABELA 7 e QUADRO 14).

Por outro lado, os resultados apontam que as instituições de vinculação dos pesquisadores possuem um processo formal estabelecido para a gestão de dados e promovem suporte técnico e boas práticas de gestão para o curto prazo, ou seja, durante o desenvolvimento do projeto. Os pesquisadores indicaram que o armazenamento de dados durante o desenvolvimento da pesquisa é fortemente satisfatório.

TABELA 7 - COMO A SUA INSTITUIÇÃO E/OU GRUPO DE PESQUISA ESTÁ ENVOLVIDA COM SEUS DADOS

Constructos	Mediana	Moda	Soma	Desvio Padrão	Variância
processo formal estabelecido para gestão de dados durante a vida do projeto (curto prazo)	3,00	4	220	1,112	1,236
processo formal estabelecido para gestão de dados além da vida do	3,00	3	200	1,106	1,223

projeto (longo prazo)					
processo para armazenar meus dados além da vida do projeto (longo prazo)	4,00	5	277	1,394	1,943
Provê ferramentas e suporte técnico para gestão de dados durante a vida do projeto (curto prazo)	4,00	4	244	1,096	1,201
Provê ferramentas e suporte técnico para gestão de dados além da vida do projeto (longo prazo)	3,00	3	215	1,129	1,275
Provê treinamentos e melhores práticas em gestão de dados	2,00	3	178	1,054	1,111
Provê recursos necessários para gestão de dados durante a vida do projeto (curto prazo)	3,00	3	199	1,017	1,014
Provê recursos necessários para gestão de dados além da vida do projeto (longo prazo)	3,00	3	200	1,014	1,029

FONTE: Dados da pesquisa (2020).

QUADRO 14 - DISTRIBUIÇÃO DE FREQUÊNCIAS - O ENVOLVIMENTO DAS INSTITUIÇÕES E/OU GRUPOS DE PESQUISA COM OS DADOS PRODUZIDOS PELOS PESQUISADORES

Questão 11	Discordo Fortemente	Discordo	Não discordo, nem concordo	Concordo	Concordo Fortemente
processo forma estabelecido para gestão de dados durante a vida do projeto (curto prazo)	8(11%)	16(21,9%)	20(27,4%)	25(43,2%)	4(5,5%)
processo formal estabelecido para gestão de dados além da vida do projeto (longo prazo)	10(13,75)	20(27,4%)	28(38,4%)	9(12,3%)	6(8,2%)
processo para armazenar meus dados além da vida do projeto (longo prazo)	8(11%)	9(12,3%)	4(5,5%)	21(28,8%)	31(42,5%)
Provê ferramentas e suporte técnico para gestão de dados durante a vida do projeto (curto prazo)	8(11%)	5(6,8%)	20(27,4%)	34(46,6%)	6(8,2%)
Provê ferramentas e suporte técnico para gestão de dados além da vida do projeto (longo prazo)	10(13,7%)	13(17,8%)	26(35,6%)	19(26%)	5(6,8%)
Provê treinamentos e melhores práticas em gestão de dados	16(21,9%)	22(30,1%)	24(32,9%)	9(12,3%)	2(2,7%)
Provê recursos necessários para gestão de dados durante a vida do projeto (curto prazo)	11(15,1%)	16(21,9%)	29(39,7%)	16(21,9%)	1(1,4%)
Provê recursos necessários para gestão de dados além da vida do projeto (longo prazo)	9(12,3%)	19(26%)	30(41,1%)	12(16,4%)	3(4,1%)

FONTE: Dados da pesquisa (2020).

A questão 13 buscou conhecer a opinião dos pesquisadores quanto ao uso de dados científicos dentro de suas áreas de pesquisa. A finalidade foi obter a

perspectiva do pesquisador sobre o uso de dados em áreas do conhecimento específicas. Os dados foram descritos na Tabela 8.

TABELA 8 - OPINIÃO DOS PESQUISADORES QUANTO AO USO DE DADOS DENTRO DE SUAS ÁREAS DE PESQUISA

Constructos	Mediana	Moda	Soma	Desvio Padrão	Variância
A falta de acesso aos dados gerados por outros pesquisadores e/ou instituições é o principal impedimento para o progresso da ciência	3,00	3	228	1,027	1,054
A falta de acesso aos dados gerados por outros pesquisadores e/ou instituições restringe a minha habilidade de responder a questionamentos	4,00	4	249	0,969	0,940
Os dados podem ser mal interpretados devido sua complexidade	3,00	3	242	0,864	0,747
Os dados podem ser mal interpretados devido a sua baixa qualidade	4,00	3	268	0,867	0,752
Os dados podem ser usados de maneiras adversas além das pretendidas	3,00	3	243	0,817	0,668

FONTE: Dados da pesquisa (2020).

Considerando que as áreas de Ciências Biológicas e Ciências Sociais Aplicadas foram indicadas como as áreas do conhecimento em que atuam a maioria dos pesquisadores respondentes, com 31,5% e 19,2%, respectivamente, conforme indicam os resultados na Tabela 4, as respostas à questão 13 indicam que essas áreas apresentam comportamento neutro quanto ao uso de dados, pois não foram percebidas dentro dessas áreas que os dados podem ser bem ou mal interpretados; que a ausência destes pode ou não restringir a habilidade de respostas dos pesquisadores a problemas da academia ou sociedade, e nem que a presença ou ausência pode impedir a resposta à questionamentos que se apresentem. Dessa forma, a partir dos resultados obtidos nesta questão, não foi possível compreender o impacto que os pesquisadores percebem no uso de dados com relação às áreas do conhecimento em que atuam.

4.2.3 Comportamento dos pesquisadores quanto ao reuso de dados

A Parte III do questionário foi composta por 12 questões cujo objetivo foi identificar o comportamento dos pesquisadores quanto ao reuso de dados, no sentido de extrair categorias capazes de identificar o impacto percebido pelo reuso de dados.

A questão 14 teve como objetivo verificar o grau de discordância e concordância dos pesquisadores sobre aspectos práticos do reuso de dados, tais como: uso de dados de outros pesquisadores, condições para disponibilização de seus dados em repositórios e compartilhamento de dados. Os resultados estão na Tabela 9.

TABELA 9 - COMPORTAMENTO DOS PESQUISADORES QUANTO A REUSO DE DADOS

Constructos	Mediana	Moda	Soma	Desvio Padrão	Variância
Eu usaria conjunto de dados de outros pesquisadores se eles fossem facilmente acessíveis	5,00	5	334	0,551	0,303
Eu estaria disposto a colocar pelo menos alguns dos meus dados em um repositório de dados sem restrições	5,00	5	323	0,912	0,831
Eu estaria disposto a colocar todos os meus dados em um repositório de dados sem restrições	4,00	5	280	1,229	1,511
Estaria mais favorável a colocar meus dados disponíveis se eu pudesse acrescentar condições de acesso	5,00	5	302	1,182	1,398
Estou satisfeito com minha habilidade em integrar dados de diferentes fontes para minhas questões de pesquisa	4,00	4	267	0,820	0,673
Estaria disposto a compartilhar dados em um amplo grupo de pesquisadores que utilizam dados de diferentes formas	5,00	5	328	0,580	0,337
É importante que meus dados sejam citados quando utilizados por outros pesquisadores	5,00	5	355	0,419	0,175
É apropriado criar novos conjuntos de dados a partir de dados compartilhados	5,00	5	336	0,618	0,382

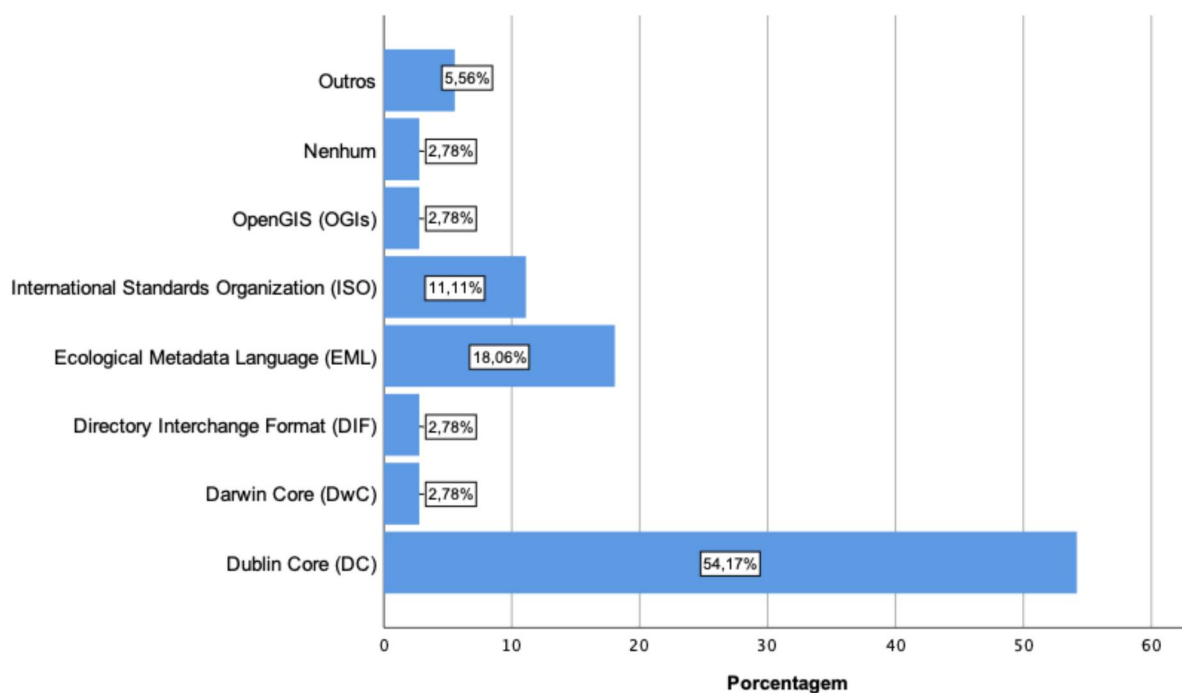
FONTE: Dados da pesquisa (2020).

A maioria dos pesquisadores concordou fortemente em reusar dados de outros pesquisadores e compartilhar seus dados em repositórios específicos sem restrições de acesso. Demonstraram também estarem satisfeitos com suas habilidades de integrar dados de diferentes fontes. O alto índice de concordância

indica que os pesquisadores se sentem confortáveis tanto para reusar dados quanto para compartilhar seus dados com outros.

A questão 15 perguntou aos pesquisadores qual padrão de metadados são utilizados para a descrição dos seus dados. Os padrões mais utilizados pelos pesquisadores foram o *Dublin Core* (DC), com 54,17% dos resultados, seguido do *Ecological Metadata Language* (EML) com 18,06%, de acordo com o Gráfico 9.

GRÁFICO 9 – PADRÕES DE METADADOS MAIS UTILIZADOS



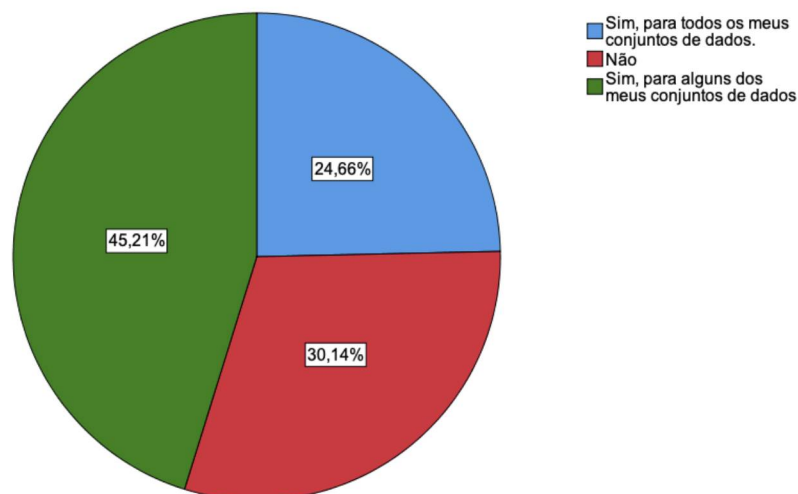
FONTE: Dados da pesquisa (2020).

Os dois padrões possuem ampla comunidade aberta de desenvolvedores e estão em permanente processo de atualização para atender às novas especificidades da produção, uso, compartilhamento e reuso de dados.

Na questão 16 foi perguntado aos pesquisadores se eles são os únicos responsáveis por aprovar o acesso aos dados, no sentido de torná-los disponíveis. Os resultados no Gráfico 10 indicaram que mais de 45% são responsáveis por dispor parcialmente seus dados para acesso; enquanto apenas 24,66% podem liberar o acesso total aos seus conjuntos de dados. Interessante observar que parte significativa dos pesquisadores não possui responsabilidade para liberação de acesso aos dados de pesquisa. Esta pergunta teve por objetivo descobrir se dentre o universo de dados produzidos pelos pesquisadores e/ou seus grupos, qual parcela

está, efetivamente, disponível para acesso e, nesse sentido, os resultados demonstraram que cerca de 30% deles não estão.

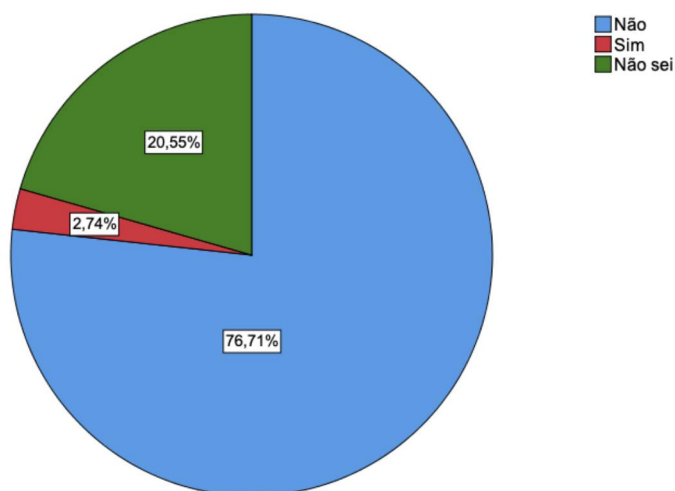
GRÁFICO 10 - ABERTURA DOS DADOS PARA ACESSO



FONTE: Dados da pesquisa (2020).

A questão 17 teve por objetivo verificar se as instituições que concedem financiamento para a realização de pesquisas solicitam a elaboração de um Plano de Gestão de Dados (PGD) aos pesquisadores. O PGD é uma ferramenta que descreve a obtenção, organização, tratamento, preservação, documentação, uso e compartilhamento de dados durante e após o projeto, facilitando a compreensão e reutilização futura (AVENTURIER, 2017). Os resultados coletados nesta pesquisa, indicados no Gráfico 11, demonstraram que a maioria das instituições financiadoras não exige PGD na concessão do recurso (76,71%) Esse resultado está de acordo com o referencial teórico no qual se apurou que no Brasil as agências, com exceção da FAPESP, ainda não exigem o planejamento de dados das pesquisas.

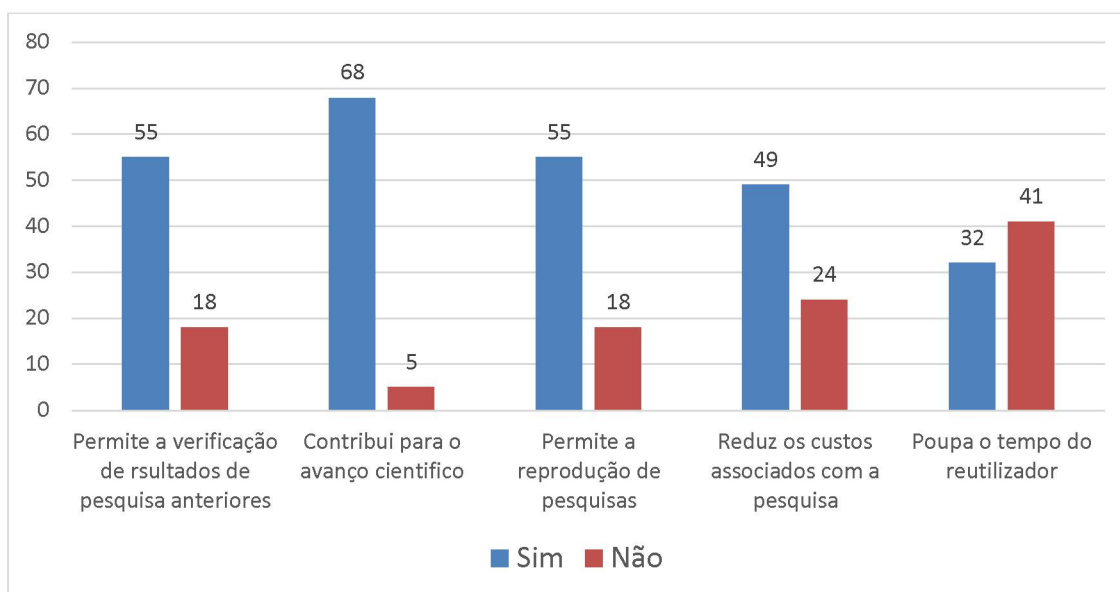
GRÁFICO 11 - EXIGÊNCIA DE PDG PELAS ENTIDADES FINANCIADORAS DE PESQUISA



FONTE: Dados da pesquisa (2020).

O objetivo da questão 18 foi saber quais benefícios os pesquisadores destacam quanto ao reuso de dados. Foram apresentados cinco constructos e os pesquisadores poderiam concordar ou não. Os resultados obtidos indicam que ampla maioria concorda com os benefícios advindos do reuso de dados, destacando-se: a contribuição para o avanço científico (68), a verificação de resultados de pesquisas anteriores (55) e a reprodução de pesquisas (55). Por outro lado, 41 pesquisadores se opuseram na questão do reuso de dados poupar o tempo do pesquisador.

GRÁFICO 12 – BENEFÍCIOS PERCEBIDOS SOBRE O REUSO DE DADOS



FONTE: Dados da pesquisa (2020).

A questão 19 mediu o grau de discordância e concordância dos constructos quanto aos riscos percebidos no reuso de dados científicos, conforme Tabela 10:

TABELA 10 - RISCOS PERCEBIDOS NO REUSO DE DADOS

Constructos	Mediana	Moda	Soma	Desvio Padrão	Variância
Receber reconhecimento inferior	1,00	1	106	0,602	0,362
Ter minha pesquisa mal considerada	1,00	1	106	0,668	0,446
Infringir involuntariamente os códigos éticos	2,00	2	142	0,880	0,775
Infringir involuntariamente as normas de proteção de dados	2,00	2	150	0,926	0,858
Não ter consentimento informado pelos participantes de uma pesquisa	2,00	2	169	0,998	0,997
Interpretar mal os dados	3,00	3	205	0,967	0,935
Fazer suposições ou associações incorretas com base nos dados	3,00	3	216	0,920	0,846
Usar dados indevidos	3,00	3	178	0,913	0,833
Fiar vulnerável a dados com erros ocultos	4,00	4	255	0,868	0,753
Ter capacidade limitada para identificar erros nos dados	3,00	4	255	0,868	0,753
Reutilizar dados imprecisos	3,00	4	243	0,987	0,974

FONTE: Dados da pesquisa (2020).

Os resultados indicam discordância sobre riscos percebidos no reuso de dados. Os pesquisadores apontam que não há riscos pessoais ao assumir o reuso de dados ao discordar, por exemplo, que o reuso possa significar a quebra de códigos éticos e normas de proteção de dados.

Os pesquisadores mostraram-se neutros quanto aos riscos de má interpretação ou suposições errôneas no reuso de dados. Todavia, uma parcela concordou que se sentem vulneráveis ou incapacitados para identificar dados com erros ocultos ou imprecisos.

A questão 20 representou a opinião dos pesquisadores sobre os esforços percebidos no reuso de dados.

TABELA 11 - ESFORÇOS PERCEBIDOS NO REUSO DE DADOS

Constructos	Mediana	Moda	Soma	Desvio Padrão	Variância
Encontrar formas inovadoras de abordar dados existentes	3,00	3	236	1,021	1,042
Justificar a importância de um novo estudo com base nos dados existentes	3,00	3	225	1,024	1,049
Identificar novos dados relevantes	4,00	5	303	1,126	1,269

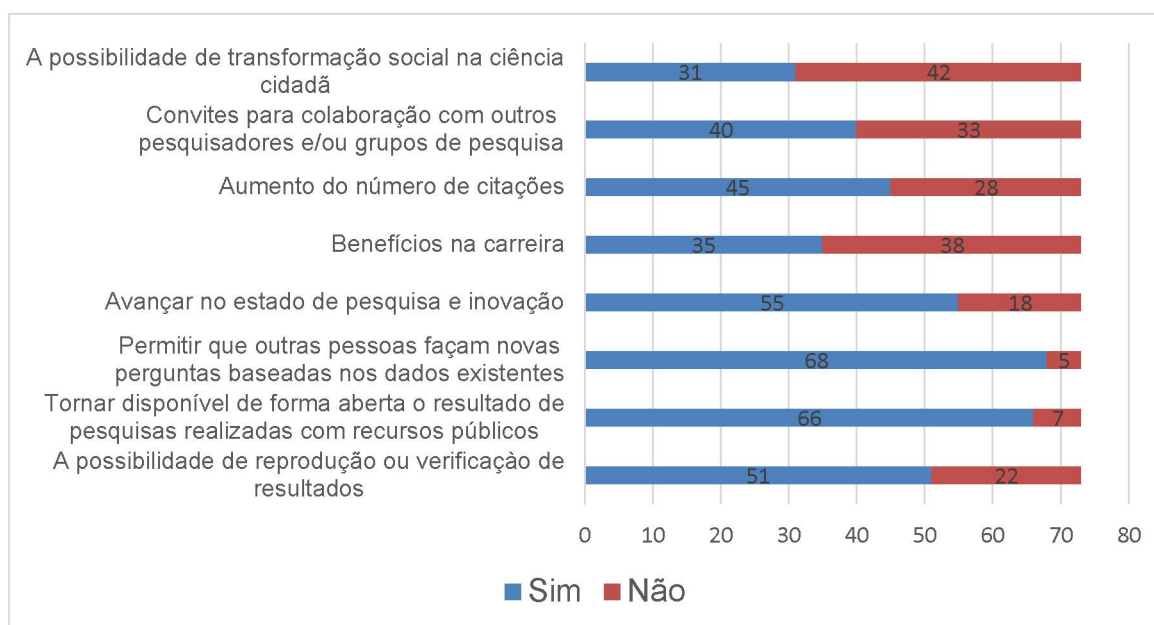
Selecionar dados reutilizáveis	5,00	5	337	0,775	0,601
Obter permissão para reutilizar dados	3,00	3	246	0,842	0,709
Recuperar dados e familiarizar-se com eles	3,00	3	257	0,784	0,614
Coletar e produzir dados por outras pessoas	4,00	4	251	0,850	0,722
Compreender o estudo original	3,00	3	220	0,890	0,791
Ajustar seu próprio projeto de pesquisa para acomodar os dados existentes	4,00	4	256	0,784	0,615
Alterar uma ideia de pesquisa para acomodar os dados existentes	4,00	4	253	0,801	0,641
Reorganizar seu estudo original para acomodar os dados existentes	4,00	4	252	0,851	0,723

FONTE: Dados da pesquisa (2020).

De acordo com a Tabela 11, a maioria dos pesquisadores demonstraram-se neutros com relação aos esforços percebidos no reuso de dados. Os resultados indicam que os processos de encontrar formas inovadoras de buscar dados, justificar importância de uso desses dados em novas pesquisas, obter permissão para reuso e familiarizar-se com eles não requerem necessariamente “esforços”. Já os processos de coleta, produção, acomodação de projeto e ajuste de ideias para incorporação de dados requerem esforço dos pesquisadores. A identificação e seleção de novos dados foi apontada com os itens que mais requerem esforços dos pesquisadores quanto ao reuso de dados.

Após verificar os riscos e benefícios que podem ser advindos do reuso de dados científicos, a questão 21 perguntou aos pesquisadores quais oportunidades eles perceberam na prática do reuso de dados. Os detalhes estão no gráfico abaixo:

GRÁFICO 13 - OPORTUNIDADES PERCEBIDAS NO REUSO DE DADOS



FONTE: Dados da pesquisa (2020).

A maioria significativa dos pesquisadores concorda que, de modo geral, com as vantagens do reuso de dados mencionados nos constructos acima. Destacam-se as afirmações: permitir que outras pessoas façam novas perguntas baseadas nos dados existentes (68 respostas), tornar disponível de forma aberta o resultado de pesquisas realizadas com recursos públicos (66 respostas) e avançar no estado de pesquisa e inovação (55 respostas). Nesse sentido, pode-se ressaltar que os pesquisadores observam que há mais vantagens no reuso de dados científicos do que riscos e esforços (TABELAS 10 e 11).

Na questão 22, perguntou-se ao pesquisador qual o grau de discordância e concordância sobre as características que os dados científicos deveriam apresentar para serem passíveis de reuso.

TABELA 12 - CARACTERÍSTICAS DOS DADOS PARA REUSO

Constructos	Mediana	Moda	Soma	Desvio Padrão	Variância
Ter documentação suficiente (por exemplo: livro de códigos)	5,00	5	332	0,708	0,501
Ter documentação simples e clara	5,00	5	345	0,449	0,202
Estar na mesma unidade de análise que eu preciso (por exemplo: indivíduos, grupos, etc)	4,00	5	306	0,995	0,990
Encaixar no meu estudo	5,00	5	314	0,967	0,936
Estar conforme o formato que eu preciso	4,00	4	280	1,067	1,139
Ser originário de uma fonte confiável	5,00	5	337	0,615	0,379

Ser o produto de uma fonte confiável	4,00	5	301	0,832	0,693
Ser consistente/estável	4,00	5	305	0,839	0,704
Estar completo ou poucos dados ausentes	5,00	5	345	0,629	0,396
Ser produto de um estudo bem projetado	3,00	3	261	0,865	0,748
Pertencer a um estudo que realizou um plano de pesquisa original	3,00	2	197	1,076	1,158
Ser originado de um estudo bem executável	3,00	3	250	0,881	0,250

FONTE: Dados da pesquisa (2020).

Ampla maioria dos pesquisadores concordou fortemente que o reuso de um dado científico requer como características o uso de documentação, e que esta seja simples e clara; que os dados devem ser completos (ou com poucas ausências); que devem estar na mesma unidade de análise utilizada pelo pesquisador; e que devem ser provenientes e produtos de fonte confiável.

As respostas indicaram também que o reuso de dados não precisa partir de um estudo original, o que fortalece a continuidade de reuso a partir de novas variações.

A questão 23 teve por objetivo conhecer em que grau os pesquisadores estão satisfeitos com suas habilidades em encontrar repositórios de dados em suas áreas de conhecimento e também verificar o nível de sua capacitação em gestão de dados. Os resultados estão na Tabela 13 a seguir:

TABELA 13 - DESCOBERTA DE REPOSITÓRIOS DE DADOS

Constructos	Mediana	Moda	Soma	Desvio Padrão	Variância
Estou ciente dos repositórios de onde posso reutilizar dados	3,00	3	255	0,884	0,781
Posso encontrar facilmente repositórios com dados relacionados à minha pesquisa	3,00	3	226	0,802	0,643
Os pesquisadores primários estão disponíveis para fornecer informações adicionais sobre seus dados	2,00	2	166	0,902	0,813
Posso facilmente entrar em contato com os pesquisadores primários para obter esclarecimentos sobre seus dados, se eu precisar	2,00	2	153	1,030	1,060
Recebi treinamento sobre como encontrar dados que eu possa potencialmente reutilizar	2,00	1	161	1,247	1,554
Tenho acesso a treinamento formal sobre habilidades úteis para reutilizar dados	2,00	1	168	1,187	1,408

FONTE: Dados da pesquisa (2020).

Os pesquisadores demonstraram neutralidade quanto a sua satisfação em encontrar repositórios de dados em suas áreas do conhecimento. Com relação a possibilidade de contato com o autor original de um conjunto de dados, as respostas mostraram discordância dos pesquisadores, ou seja, os pesquisadores foram quase unânimes sobre a indisponibilidade de contato com o autor original de um *dataset*. Fato esse que pode ser comprovado por esta pesquisa na dificuldade de encontrar o contato dos pesquisadores junto aos metadados de autoria nos conjuntos de dados.

Em relação ao recebimento de capacitação para encontrar e reusar dados, os resultados obtidos indicaram que os pesquisadores discordam fortemente que tenham acesso a capacitações e treinamentos. Apenas 17,8% dos pesquisadores receberam capacitação formal para gestão de dados e somente 16,4% responderam que têm à sua disposição acesso a treinamento formal em dados. Essa informação explicita a necessidade pungente de letramento em dados para reuso de dados científicos de pesquisa, sendo esse um nicho importante nas atribuições educacionais das bibliotecas universitárias e de centros de pesquisa.

A questão 24 perguntou o grau de satisfação dos pesquisadores sobre o reuso de dados científicos dentro de suas áreas de pesquisa. Conforme a Tabela 14, os resultados indicam que os pesquisadores concordam que suas áreas são receptivas quanto ao reuso de dados. Considerando que a maioria dos respondentes indicou Ciências Biológicas e Ciências Sociais Aplicadas como áreas do conhecimento de suas pesquisas, nota-se que há discrepância entre a percepção de aceitação de reuso em relação ao reuso de forma efetiva, pois as respostas apontam comportamento neutro quanto ao reuso como prática comum. Por outro lado, as respostas indicam que quase 77% dos discordam que suas áreas não têm interesse no reuso de dados.

TABELA 14 - OPINIÃO SOBRE REUSO DE DADOS DENTRO DAS ÁREAS DE PESQUISA

Constructos	Mediana	Moda	Soma	Desvio Padrão	Variância
É receptiva quanto a reuso de dados	4,00	4	264	0,892	0,795
Tem o reuso de dados como prática comum	3,00	3	219	0,882	0,778
Convenientemente reusa dados	3,00	3	218	0,890	0,791
Não tem interesse em reusar dados	2,00	2	143	0,716	,0512

FONTE: Dados da pesquisa (2020).

A última questão quis saber a opinião dos pesquisadores sobre o apoio ao reuso de dados. Os resultados obtidos apontam unanimidade quanto ao incentivo próprio ao reuso, apoio à causa e ao efetivo reuso de dados.

TABELA 15 - OPINIÃO SOBRE REUSO DE DADOS

Constructos	Mediana	Moda	Soma	Desvio Padrão	Variância
Incentivar-me a reusar dados	5,00	5	341	0,502	0,252
Apoiar o reuso de dados	5,00	5	349	0,449	0,201
Reusar dados frequentemente	5,00	5	333	0,645	0,416

FONTE: Dados da pesquisa (2020).

As 25 questões acima amparadas por 103 constructos apresentaram o perfil e os comportamentos de uso e reuso de dados por parte de 73 pesquisadores que responderam a pesquisa. Nota-se que o comportamento dos pesquisadores é aderente ao uso, compartilhamento e reuso de dados. A análise apurada dos resultados será apresentada na seção a seguir.

4.3 CATEGORIZAÇÃO DO IMPACTO PERCEBIDO PELOS PESQUISADORES

O objetivo específico “d” estabelecido nesta pesquisa foi categorizar os fatores de impacto percebido pelos pesquisadores quanto ao reuso de dados científicos. Para tanto, foi necessário estabelecer: a) o conceito de impacto dentro de uma abordagem comportamentalista; e b) quais são as atitudes que levam os pesquisadores ao reuso de dados a partir das respostas colhidas no questionário.

Para definição do conceito de impacto foi utilizada a Teoria da Ação Racional (TRA) proposta por Fishbein e Ajzen no final da década de 1960 a partir de estudos sobre atitude e comportamento. A TRA é uma teoria advinda da psicologia social para o estudo das disposições comportamentais humanas a partir da definição e relacionamento entre atitude, crença, normas subjetivas e comportamento (FISHBEIN; AJZEN, 1974).

A TRA foi inicialmente aplicada ao campo da psicologia para a predição de comportamento social a fim de compreender a escolha humana em situações diversas, como por exemplo: compartilhar ou não dados científicos; reusar ou não reusar dados de outros pesquisadores. A teoria teve ampla utilização no campo da saúde, educação, prática clínica, estudos na área de *marketing*, comportamento do

consumidor e diversas áreas dentro das ciências sociais aplicadas (MOUNTINHO; ROAZZI, 2010).

No modelo unidimensional proposto por Fishbein e Ajzen quanto ao conceito de atitude, no entendimento de Lima e D'amorim (1986, p. 135), as atitudes são conceituadas como uma "predisposição aprendida para responder de maneira consistentemente favorável ou desfavorável a respeito de um dado objeto". Ou ainda, considera "atitude como a qualidade de afeto pró ou contra algum objeto". Abaixo, adaptou-se o exemplo utilizado por Lima e D'amorim (1986) para compreender melhor o significado de atitude.

Reusar dado é

Econômico.....Dispendioso

Onde, "reusar dado" é o objeto e "econômico" ou "dispendioso" são atributos. De modo análogo, seria possível dizer que a atitude do pesquisador em reusar dados causa um efeito que pode ser econômico ou dispendioso, por exemplo. Deste modo, tem-se então que a atitude pode ser definida em função de uma crença, estímulo ou qualidade que leva indivíduo a acreditar/agir em determinado sentido.

A considerar impacto como um efeito transformador em um objeto ou situação decorrente de uma tomada de atitude (VILLAVECES et al., 2005), é possível estabelecer um paralelo de que uma atitude tomada por um indivíduo está baseada em um impacto que ele visa gerar com seu comportamento. Assim, é possível estender esse entendimento e compreender que a atitude que um pesquisador toma ao reusar dados pode ser considerada também como um fator de impacto.

Para Villaveces et al. (2005, p. 126, tradução nossa), "a análise de impacto deve ter em conta as consequências buscadas, as atividades programadas e realizadas com os meios adequados, os resultados produzidos por tais atividades, e a relação entre estes e as intenções declaradas", o que vai de encontro com Fishbein e Ajzen (1975) de que o comportamento de um indivíduo é baseado em suas atitudes (dentre outros aspectos) e que estas têm por objetivo produzir um efeito, uma modificação na situação atual, ou seja, causar um impacto.

Dessa forma, ações como: definição da tipologia do dado, padrão de descrição, preparação de dados, codificação, elaboração de documentação,

licenciamento, depósito em repositório, estabelecimento de critérios de reuso, entre outros, são atitudes tomadas por pesquisadores em busca de um resultado (reuso, visibilidade, citação, etc.) capaz de gerar impacto (benefício, oportunidade, etc.).

Definido o conceito de impacto (a), o ato seguinte foi elaborar as categorias de impacto a partir das atitudes (b) tomadas pelos pesquisadores por meio das respostas às partes II e III questionário. Foram selecionadas as respostas que tivessem relação direta com o reuso de dados.

Deste modo, foram levantados 22 fatores de impacto quanto ao reuso. Os fatores foram exportados para o *software Atlas.ti* (versão 8.4.4), sendo codificados e criadas cinco categorias, descritas no quadro abaixo.

QUADRO 15 - CATEGORIAS E FATORES DE IMPACTO

Fatores	Categorias de Impacto
Procedimentos de coleta de dados	Condições facilitadoras
Processos de análise dados	
Suporte para armazenamento do projeto (durante e após a pesquisa)	
Ferramentas para análise de dados	
Ferramentas para criação de dados	
Suporte a metadados	
Verificação de pesquisas	Benefícios
Transparência no investimento de recursos públicos	
Reprodução de pesquisas	
Aumento de citações	
Ausência de dados restringe a habilidade de responder a questionamentos	Riscos
Vulnerabilidade a dados com erros ocultos	
Capacidade limitada para encontrar erros nos dados	
Utilizar dados imprecisos	
Identificar novos dados relevantes	Esforços
Selecionar dados reutilizáveis	
Convite para colaboração	Oportunidades
Avançar no estado de pesquisa e inovação	
Tornar abertos os dados e resultados de pesquisas financiadas com recursos públicos	

FONTE: Dados da pesquisa (2020).

A categoria “Condições facilitadoras” concentra os fatores de impacto que são condicionantes estruturais, técnicas e de suporte para adoção do reuso de dados.

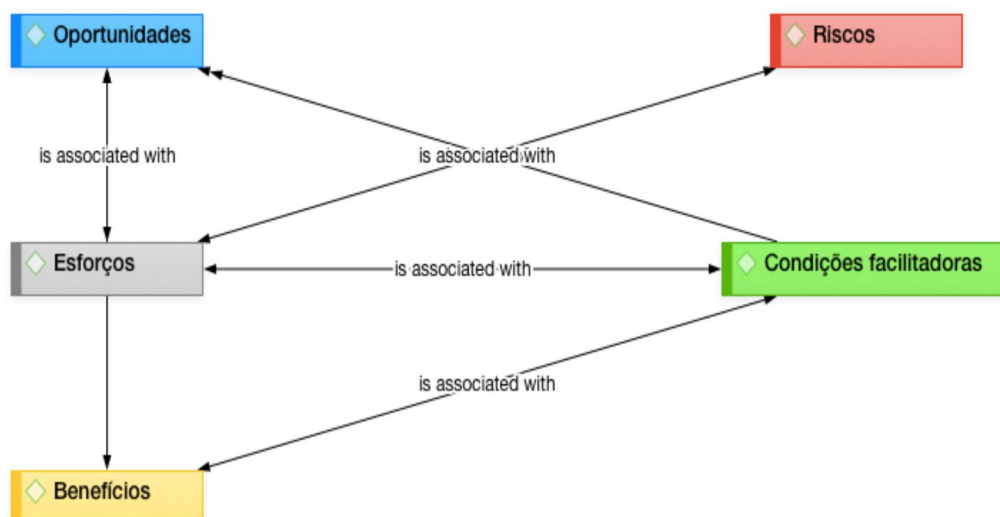
As categorias “Benefícios” e “Oportunidades”, embora contenham fatores de impacto que podem ser considerados próximos, no sentido de conferirem vantagem ao reuso de dados, apresentam fatores de benefícios e oportunidades tanto pessoais para os pesquisadores, quanto institucionais. A tomar como exemplo o fator “Convite para colaboração”, este pode assumir tanto um caráter de oportunidade profissional ao pesquisador e também de benefício pessoal ao assumir que a colaboração científica pode acarretar vantagens curriculares e de experiência em pesquisa.

A categoria “Esforços” consiste em dois fatores relevantes sobre o reuso de dados que estão diretamente relacionados a atividade básica metodológica de seleção e coleta de dados científicos, que envolve aspectos como localização de repositórios, estratégia de busca, seleção, filtragem, coleta, análise, descarte e uso.

A categoria “Riscos” envolve os fatores de impacto relacionados com a qualidade de um conjunto de dados, principalmente no quesito de completude e erros ocultos. Ainda que o conjunto de dados seja produto e/ou tenha proveniência a partir de uma fonte confiável, da feita que a avaliação das pesquisas é realizada, em sua maioria, com base no relatório (artigo) e não nos dados que geraram esse relatório, os dados podem estar comprometidos com erros ocultos percebidos tardiamente.

As cinco categorias identificadas apresentam relações de interdependência, conforme o desenho da rede na Figura 15, sendo a categoria “Condições facilitadoras” a conjuntura basilar para a adoção ou não do reuso de dados.

FIGURA 15 - REDE DE CATEGORIAS DE IMPACTO



FONTE: Dados da pesquisa (2020).

A categorização dos fatores de impacto percebidos quanto ao reuso de dados nesta pesquisa teve por objetivo identificar as atitudes que impactam no comportamento dos pesquisadores quanto ao reuso, levantando desafios, dificuldades, benefícios, riscos ou vantagens, e não se objetivou medir quantitativamente a extensão do impacto no sentido psicométrico da atividade científica de cada pesquisador individual, pois seria necessário volume de tempo para adentrar nesse nível de detalhamento.

A literatura sobre reuso de dados científicos aborda a questão do impacto sob a ótica estrita da quantificação de resultados para fins de visibilidade e citação (BELTER, 2014; NIELSEN; HJØRLAND, 2014; BALL; DUKE, 2015; FERRER-SAPENA et al. 2016). Entretanto, essa abordagem não se demonstra totalmente adequada para medir o quanto cada atitude tomada pelos pesquisadores causa efeito sob a realidade ou atividade científica em torno do reuso de dados.

A quantificação dos resultados da produção científica faz parte do rol dos processos de comunicação da ciência. Medir o alcance e impacto de publicações faz parte dos chamados “estudos métricos” da informação, um “interdomínio” da Ciência da Informação constituído por diferentes subcampos como a Bibliometria, Cientometria, Webometria, Patentomia, Altimetria e Informetria (GRÁCIO; OLIVEIRA, 2016; FREITAS; BUFREM; GRÁCIO, 2017). A mensuração do impacto de dados científicos não poderia considerar somente os padrões métricos para sua avaliação,

visto que o impacto também está relacionado a subjetividade do comportamento do pesquisador, o que demanda estudos aprofundados para estabelecimento de um modelo avaliativo.

4.4 SÍNTESE DOS RESULTADOS

Com relação a identificação e descrição dos repositórios de dados exclusivamente brasileiros (objetivo específico “b”) foi avaliado seu atendimento quanto às seguintes características levantadas na literatura: características essenciais aos repositórios de dados; contemplação do alinhamento do repositório em relação aos critérios da *Panton Principles*; e atendimento aos requisitos de disponibilidade de informações do Re3data.org.

Posto isto, verificou-se que os repositórios atendem parcialmente às características essenciais dos repositórios quanto a armazenamento, descrição e apresentação dos itens, visto que os três repositórios não apresentaram metadados de preservação, ainda que suas políticas de dados indicassem algumas ações de preservação tal como *backup*, replicação (BDC/UFPR e PPBio, respectivamente). No caso do IBICT, o próprio repositório já é um ambiente de preservação para os dados de projetos paralelos do Instituto. Ainda sim, por sua configuração de repositório, deveria existir uma política de dados para prever medidas de salvaguarda dos seus dados.

Em relação ao alinhamento dos repositórios de dados com a *Panton Principles*, o resultado foi parcial para a BDC/UFPR e IBICT *Dataverse Network*, pois não foram plenamente atendidos os critérios de licença específica para dados, permissões de reuso, bem como esses repositórios apresentaram limitações em alguns dos seus dados para uso comercial e derivativo. O PPBio não atendeu a nenhum dos critérios da *Panton Principles*.

A respeito dos requisitos de disponibilidade (informação, acesso, licenças, identificador persistente, certificados e padrões e política), BDC/UFPR e PPBio atenderam plenamente aos itens listados quanto a informação; enquanto o IBICT atendeu parcialmente, considerando a ausência de licença individual para dados e a falta de uma política e/ou diretriz de gestão.

Em adendo especial sobre os *software* de gestão de dados, a descrição dos repositórios demonstrou como estão estruturadas as plataformas onde os dados

científicos são depositados e foi interessante destacar que os dados científicos são gerenciados a partir de soluções *open source*, demonstrando alinhamento às propostas dos movimentos da ciência aberta.

Os três *software* livre utilizados pelos repositórios, DSpace, Dataverse e MetaCat, respectivamente, apresentam um bom desempenho no suporte a gestão de dados científicos de pesquisa, ainda que a versão do DSpace utilizado pela BDC/UFPR não seja especificamente voltada para este fim, dado que a aplicação é comumente utilizada para construção de repositórios de publicações e produção científica. A extensão DSpace-CRIS³⁴ se alinhada com a versão 6.x³⁵ do *software* permitirá ao repositório melhor ambiente para o *workflow* com dados científicos, agregando também o versionamento dos projetos e *datasets* depositados.

Os *software* Dataverse e MetaCat, por sua vez, apresentam especificações customizadas para o trabalho com dados, tendo inclusive o suporte de ferramentas abertas disponibilizadas pelo *DataONE*. O *Dataverse* tem relevante adesão internacional, com mais de 435 mil³⁶ dados depositados no somatório de suas instâncias em produção.

Em relação ao licenciamento dos dados nos repositórios, constatou-se que em nenhum deles os dados são integralmente abertos permitindo ações de modificações, compartilhamento e reuso.

Em resumo, concluindo a descrição dos repositórios, as fragilidades observadas residem no tipo de licenciamento atribuído aos dados, no aspecto de preservação, e na sustentabilidade dos projetos, haja vista a indisponibilidade de acesso aos repositórios PPBio e IBICT *Dataverse* que permanecem fora do ar e o suporte por *e-mail* sem retorno dos gestores após dois meses.

A análise da descrição dos repositórios indica coerência com os requisitos levantados na revisão de literatura, pois, de fato, os repositórios estão buscando adequarem-se a padrões internacionais tanto em relação às plataformas (RÜCKNAGEL et al., 2015; SAYÃO; LEITE, 2016) quanto a padrões de metadados (SAYÃO; LEITE, 2016). Os requisitos destacados a partir de Rodrigues et al.

³⁴ DSpace-cris. Disponível em: https://dspace-cris.4science.it/?locale=pt_BR. Acesso em: 07 mar. 2020.

³⁵ DSpace 6.x. <https://wiki.duraspace.org/display/DSDOC6x/Item+Level+Versioning>. Acesso em: 07 mar. 2020

³⁶ Conforme dados do sítio eletrônico. Disponível em: <https://dataverse.harvard.edu>. Acesso em: 07 mar. 2020.

(2010) e Paganine e Costa (2016) sobre metadados estão a ser atendidos, ainda que parcialmente; bem como o licenciamento dos dados de acordo com Murray-Rust (2010) e ao padrão do RE3data.org descrito Pampel et al. (2013) e Vierkant et al. (2014).

Com relação a identificação dos pesquisadores e elaboração das categorias de fatores de impacto (objetivos específicos “c” e “d”), a parte I do instrumento de coleta de dados teve por objetivo traçar um perfil dos respondentes da pesquisa. Assim, obteve-se que os respondentes se caracterizam por exercer atuação profissional como docente em instituições acadêmicas públicas e possuem domicílio profissional na Região Sul do país. O nível de titulação desses docentes é predominantemente entre doutorado e pós-doutorado. O financiamento dos projetos de pesquisa depende quase que majoritariamente de financiamento institucional. Com relação a área do conhecimento, os pesquisadores concentram-se principalmente nas áreas de Ciências Biológicas e Ciências Sociais Aplicadas, o que está de acordo com os repositórios que serviram de ambiente para a pesquisa (dois multidisciplinares e um focado em ciências biológicas/ambientais).

Sobre o perfil dos pesquisadores, observa-se que não fora perguntado qual o nome da instituição de vinculação profissional para evitar uma possível personificação dos respondentes através do cruzamento de dados a partir da resposta de outras questões.

Conhecer o perfil dos pesquisadores foi uma introdução foi um objetivo específico necessário e também uma introdução para a análise da parte II do questionário, cujo objetivo era compreender o comportamento dos pesquisadores sobre o uso de dados científicos. Os pesquisadores indicaram que os tipos de dados mais utilizados por eles enquadram-se na categoria de dados bióticos (conjunto dos seres vivos que compõem a comunidade biológica), seguido de dados em pesquisa social, o que demonstra nova compatibilidade com os repositórios ambientes da pesquisa.

No tocante a repositórios, mais de 75% dos pesquisadores indicaram que todos os seus dados estão disponíveis em repositórios da instituição ao qual são vinculados e em repositórios de redes nacionais, o que pode ser considerado uma boa adesão ao compartilhamento. Entretanto, as causas observadas para a parcela indisponível residem em três circunstâncias, sendo a 1º: na falta de tempo dos pesquisadores para a preparação dos dados; 2º: o fato dos respondentes não serem

únicos responsáveis pela disponibilização dos seus dados; e por fim, em 3º: por não haver exigência do financiador da pesquisa para disponibilização. A primeira circunstância alerta para um ponto sensível nas atribuições da carreira docente, que reside na multiplicidade de tarefas sem o suporte adequado, em alguns casos, para execução de todas as funções.

Ainda a respeito dos repositórios, os resultados apontam que os pesquisadores estão satisfeitos com o envolvimento de suas instituições em relação ao armazenamento, suporte e boas práticas durante o desenvolvimento dos projetos de pesquisa – satisfação no curto prazo. Entretanto, apontam comportamento institucional neutro em relação às mesmas funções a longo prazo. Essa observação merece atenção por parte das bibliotecas e centros de suporte à pesquisa das instituições, pois a mensuração e impacto do uso e reuso de dados é observado também a médio e longo prazo, haja vista os estudos métricos considerarem períodos de anos para a elaboração das métricas.

A parte III do questionário objetivou compreender o comportamento dos pesquisadores especificamente quanto ao reuso de dados. Os resultados apontam que os pesquisadores concordam fortemente em reusar e compartilhar seus dados, ainda que expressaram sentir-se mais confortáveis se pudessem acrescentar condições de acesso e reuso.

Com relação a oportunidades *versus* riscos percebidos quanto ao reuso de dados, os pesquisadores indicaram que há mais benefícios do que riscos, sendo os benefícios “tornar disponível de forma aberta o resultado de pesquisas financiadas com recursos públicos” e “permitir que pessoas façam novas perguntas a partir de dados existentes” as maiores expressões de benefício. Por outro lado, dados com erros ocultos ou imprecisos são os riscos indicados com mais frequência pelos pesquisadores.

No tocante a benefícios e esforços quanto ao reuso, a oportunidade mais frequente indicada nas respostas foi contribuir para o avanço científico. Os esforços quanto ao reuso destacam-se nos aspectos de identificar e selecionar novos dados reutilizáveis, aspecto este que está relacionado com o fato de os pesquisadores não concordarem com a proposição de que o reuso poupa tempo de trabalho do pesquisador. Esse resultado faz sentido dado que o processo de seleção de dados engloba atividades como verificação de completude dos dados, qualidade e

autenticidade da fonte dos dados e testes de adequação. Um resumo dos fatores de impacto segue no Quadro 16.

QUADRO 16 - RESUMO DOS RESULTADOS OBTIDOS SOBRE AS CATEGORIAS DE IMPACTO

Condições Facilitadoras	Benefícios	Oportunidades	Esforços	Riscos
Fatores de impacto condicionantes para adoção ou não do reuso.	Assume caráter de impacto positivo, dado que os fatores levantados atuam em favor do pesquisador, possibilitando vantagem ou ganho em sua atividade científica.		Energia dispendida pelo pesquisador para encontrar novos dados relevantes e reutilizáveis (acesso e compatibilidade).	Relacionado a qualidade e integridade dos dados e na conveniência de adequar um projeto ou ideia original para acoplar dados de terceiros.

FONTE: Dados da pesquisa (2020).

De modo geral, os resultados indicaram que os pesquisadores identificam benefícios e oportunidades sobre o reuso de dados, e mostram-se fortemente engajados com as atitudes de reuso.

As cinco categorias de impacto descritas nesta seção, ainda que não exaustivas, podem servir de base para explicitar os principais fatores que afetam o comportamento dos pesquisadores quanto ao reuso de dados. Usando o preceito de atitude da TRA sobre o comportamento dos indivíduos e assumindo que impacto é “efeito que, por sua força, impede ou acarreta mudanças” (DICIONÁRIO MICHAELIS, *on-line*)³⁷; ou ainda, que impacto é “efeito ou influência que um agente, evento ou recurso tem sobre o outro” (BALL; DUKE, 2015), observou-se que as atitudes tomadas pelos pesquisadores (preparação de dados, atribuição de licenciamento aberto, identificação de novos dados relevantes, seleção de dados reutilizáveis e compartilhamento) tem como base o resultado (reuso) a partir da sua decisão sobre uma crença (o reuso de dados reduz custos de pesquisa na fase de coleta, por exemplo). Assim, é possível dizer que o comportamento do pesquisador em reusar dados é baseado nas suas crenças sobre algum tipo de benefício ou oportunidade que tenha impacto em sua atividade científica ou pessoal. Ou ainda, que o comportamento em não reusar dados pode estar baseado nos riscos percebidos de mau uso ou na incompletude dos dados compartilhados, por exemplo.

³⁷ <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/impacto>. Acesso em: 25 mar. 2020.

5 CONCLUSÃO

Esta pesquisa teve por objetivo identificar os fatores de impacto percebido pelos pesquisadores quanto ao reuso de dados científicos de pesquisa no contexto dos repositórios de dados brasileiros. Para atingir esse fim, buscou-se compreender quem são esses pesquisadores e como compartilham e reusam dados. Assim, foram elencados a partir do referencial teórico os aspectos relevantes que dão fundamento a gestão e reuso de dados.

Os repositórios de dados foram descritos a partir de um protocolo que verificou o atendimento destes às características essenciais aos ambientes quanto a: armazenamento e descrição de itens; alinhamento em relação aos critérios da *Panton Principles* (abertura dados); e atendimento aos requisitos disponibilidade de informações do Re3data.org (quanto a licenciamento aberto). O resultado dessa descrição comprovou que os repositórios brasileiros ainda não possuem toda a infraestrutura necessária para a gestão de dados científicos, pois há aspectos como licenciamento específico para dados que ainda não estão adequados às recomendações de abertura. Os repositórios carecem de serviço efetivo de curadoria de dados que seja capaz de tratar da indexação dos *datasets* em mecanismos de busca, sistemas de citação, referência e análises sobre reuso. Mas, principalmente, tratar questões ainda elementares como oferecer serviços de conversão de dados para formatos efetivamente abertos.

A partir da descrição dos repositórios também foi possível realizar a identificação dos pesquisadores que responderiam ao questionário cujas respostas ofereceriam subsídio para identificação dos fatores de impacto.

Os resultados obtidos no questionário traçaram o perfil dos pesquisadores e sua relação com os dados científicos. Da mesma forma, os resultados também avaliaram o grau de discordância ou concordância com relação aos preceitos de compartilhamento e reuso de dados, indicando concordância com as práticas da ciência aberta e atitudes de aderência e incentivo ao reuso de dados. Nesse aspecto, a satisfação dos pesquisadores foi mostrada ao destacarem que o compartilhamento e reuso de dados proporcionam um benefício a medida que é capaz de promover transparência no investimento de recursos públicos na pesquisa, bem como a possibilidade de novas análises e olhares em contextos diversos a partir da abertura de dados.

Os resultados obtidos na aplicação do questionário alcançaram os objetivos propostos na pesquisa e responderam também ao questionamento proposto nesta investigação, pois foi possível identificar 22 fatores de impacto percebidos pelos pesquisadores quanto ao reuso. Essa identificação foi extraída a partir da análise das respostas sobre o comportamento dos pesquisadores medida a partir nas escalas de discordância ou concordância com os preceitos de gestão, uso, compartilhamento e reuso de dados.

Na Ciência da Informação, o termo “impacto” é muitas vezes associado com o viés da contabilização de citações às publicações (periódicos), o que leva à geração de um índice, que se transforma em uma nota (Fator de Impacto – FI). Essa nota confere um atributo qualitativo a uma publicação, o que representa dizer, em síntese, que uma publicação é mais importante que outra. E ainda que esse atributo qualitativo seja controverso e polêmico, essa avaliação de publicações pelo FI pode ser válida para medir impacto de publicações, desde que considerada as suas limitações. Todavia, quando o fator comportamental é considerado para fins de impacto, a métrica avaliativa do FI não deveria ser única a ser considerada para mensurar impacto, sobre reuso de dados por exemplo. As atitudes que afetam o comportamento dos indivíduos não devem ser medidas apenas sob o aspecto da métrica, tal como as publicações. Antes, deve-se levar em consideração a subjetividade do comportamento humano. Logo, a medição do impacto sobre comportamento dos pesquisadores em relação ao reuso de dados não é um resultado preciso, mas sim uma tentativa de aproximação.

Uma categorização desses fatores sintetizou os impactos percebidos e tornou simplificada a compreensão de que as condições facilitadoras a nível estrutural (ferramentas, suporte, infraestrutura tecnológica, treinamento, suporte a curto e longo prazo e financiamento) são o primeiro passo para oferecer condições de disponibilidade, compartilhamento e reuso de dados.

As categorias de impacto que aparecem em sequência: benefícios, riscos, esforços e oportunidades são consequências diretas das condições facilitadoras. Por exemplo: um pesquisador que não tem suporte institucional quanto a tempo dedicado à pesquisa, ferramentas e sistemas adequados para criação/coleta, tratamento, codificação e disponibilização de dados e financiamento apropriado dificilmente terá motivação para disponibilizar seus dados e assim, não passará por

riscos e nem esforços, contudo, não poderá também usufruir de benefícios e oportunidades advindas do reuso.

A consideração acima reforça a importância da estruturação dos repositórios e o oferecimento de curadoria de dados para os pesquisadores. Os serviços de curadoria envolvem não apenas o depósito de dados, mas algo amplo, como um serviço de auxílio à pesquisa (prática comum em universidades estrangeiras) que oriente os pesquisadores a planejar todos os aspectos do seu projeto, partindo da pergunta de pesquisa até o compartilhamento de dados e atualização destes em repositórios.

Nesse aspecto, as bibliotecas universitárias, por já estarem familiarizadas com o tratamento de dados e as atividades do serviço de referência oferecido por bibliotecários, podem tomar a liderança dessa curadoria ofertando os seguintes serviços: suporte à elaboração do Plano de Gestão de Dados, seleção de repositórios de dados confiáveis, auxílio quanto a definição de padrão de dados conforme o tipo de dado a ser utilizado, auxílio à submissão de dados em repositórios, auxílio quanto ao licenciamento e demais aspectos sobre direitos autorais, auxílio no compartilhamento e no reuso de dados.

Com os achados desta pesquisa, é possível apontar uma contribuição teórica e prática sobre o compartilhamento e impacto sobre reuso de dados. Como contribuição teórica, a identificação de fatores de impacto sobre o reuso de dados numa abordagem comportamental. Ainda que esta pesquisa não tenha tido a pretensão de esgotar o assunto, é um pequeno passo para preencher a lacuna de estudos sobre impacto em reuso de dados na perspectiva brasileira e latina.

Como contribuição prática, a descrição e exploração dos repositórios de dados na busca pela identificação dos pesquisadores, permitiu verificar que os repositórios brasileiros não ainda não estão integralmente adaptados para divulgação, compartilhamento e reuso de dados. Alguns preceitos fundamentais como adesão a licenciamento específico para dados não foram implementados. É preciso atenção para questões como acessibilidade nos dados (há itens em formato PDF, por exemplo), ausência de ferramentas de exportação, citação e definição de ações mais específicas e efetivas de preservação digital. Essas observações podem ser úteis para os gestores de repositórios na elaboração de política de dados.

5.1 DIFICULDADES NA PESQUISA E SUGESTÕES DE PESQUISAS FUTURAS

Uma dificuldade para obtenção dos dados da pesquisa ocorreu devido a indisponibilidade na rede de dois dos três repositórios que foram ambiente de análise, isso no momento da coleta de dados, o que atrapalhou o cronograma da pesquisa. A análise do repositório PPBio Data Repository só foi possível devido a sua estratégia de preservação digital em replicar os dados e o ambiente *ipis literis* para outra plataforma. O repositório IBICT Dataverse Network continua indisponível e a pesquisadora não obteve retorno dos contatos via e-mail para a administração do serviço. A análise deste repositório na pesquisa só foi possível devido uma coleta preliminar que antecedeu a qualificação do projeto de dissertação.

Como sugestão de pesquisa futura, realizar a validação dos fatores de impacto extraídos das respostas no questionário pode ser uma forma de testar o modelo e comprovar sua aplicabilidade.

Outra sugestão de pesquisa futura, destaca-se a continuação dos estudos sobre impacto em reuso de dados numa abordagem comportamental que explore psicometricamente o impacto do reuso causado nos pesquisadores na área de Ciências Sociais Aplicadas.

Os dados obtidos nesta pesquisa estarão receberão licenciamento CC0 e estarão disponíveis para reuso na Base de Dados Científicos da UFPR.

REFERÊNCIAS

AGRESTI, A.; FINLAY, Barbara. **Métodos estatísticos para as ciências sociais**. 4.ed. São Paulo: Penso, 2012.

ARAÚJO, L. M.S.; MÁRDERO ARELLANO, M.A.; FERRER, I. D. **Guia para os usuários do repositório Dataverse IBICT**. Brasília: IBICT, 2018. Disponível em: http://cariniana.ibict.br/images/artigos/Dataverse/Guia_Usuarios_Dataverse_Ibict.pdf. Acesso em 17 dez. 2019.

ARAÚJO, P.C. **Contribuições da gestão da informação para o sub-processo de coleta do processo de inteligência competitiva nas empresas prestadoras desse serviço no sul do Brasil**. 2011. Dissertação (Mestrado em Ciência, Gestão e Tecnologia), Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, Curitiba, 2011. Disponível em: <http://hdl.handle.net/1884/30894>. Acesso em: 20 set. 2019.

AVENTURIER, P. **Princípios FAIR**: critérios de qualidade para dados de pesquisa. Disponível em: <https://publicient.hypotheses.org/1456>. Acesso em: 12 set. 2019.

AZEVEDO, D. J. S. et al. Biotic or abiotic factors: wich has greater influence in determing the structure of rotifers in semi-arid reservoirs? **Acta Limnologica Brasiliensia**, v. 17, n. 1, 2015. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2179-975X2015000100007. Acesso em: 18 set. 2019.

ALBAGLI, S.; MACIEL, M.L.; ABDO, A.H. (Orgs). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. Disponível em: http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20%285%29.pdf. Acesso em: 22 ago. 2018.

BALL, A.; DUKE, M. How to track de impact of research data with metrics. **DDC How- to-guides**. Edinburg: Digital Curation Center, 2015. Disponível em: <https://www.dcc.ac.uk/guidance/how-guides/track-data-impact-metrics#impact-measurement-concepts>. Acesso em: 17 mar. 2020.

BELTER, C. W. Measuring the value of research data: a citation analysis of oceanographic data sets. **PLoS ONE**, v. 9, n. 3, 2014. Disponível em: <https://doi.org/article/6aec4d4bc7334d8f83ea39b0f896c558>. Acesso em: 12 mar. 2020.

BORGMAN, C. L. The conundrum of sharing research data. **Journal of The American Society for Information Science and Technology**, v. 63, n. 3, p. 1059-1078, 2012. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.22634>. Acesso em: 07 ago. 2018.

BUENO DE LA FUENTE, G. Challenges and strategies for the success of Open Science. In: **FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN**

RESEARCH (FOSTER). Open Science at the Core of Libraries. 2016. Disponível em: <https://www.fosteropenscience.eu/node/1424>. Acesso em 07 ago. 2018.

BUFREM, L.S. et al. Produção internacional sobre ciência orientada a dados: análise dos termos data Science e e-science na Scopus e na Web of Science.

Informação & Informação, Londrina, v.21, n.2, p.40-67, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/26543>. Acesso em: 12 dez. 2018.

CALLON, M. Society in the Making: The Study of Technology as a Tool for Sociological Analysis. IN: In: PINCH, Trevor J.; BIJKER, Wiebe E.; HUGHES, Thomas P. **The social Construction of Technological Systems**. Cambridge (Mass.), MIT, 1993, p. 77-97.

CAMPOS, L. F. B. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Encontros Bibli: Revista Eletrônica Biblioteconomia e Ciência da Informação**, Florianópolis, n. 23, 1, p. 16-46, 2007. Disponível em: <https://www.brapci.inf.br/index.php/article/download/11673>. Acesso em: 02 fev. 2020.

CATLIN, A. C. et al. Ac cyberplatform for sharing scientific research data at DataCenterHub. **Computing in Science & Engineering**, v 20, n. 49, 2018. Disponível em: <https://aip.scitation.org/doi/abs/10.1109/MCSE.2017.3301213>. Acesso em: 30 jan. 2019.

CONSELHO NACIONAL DAS FUNDAÇÕES ESTADUAIS DE AMPARO À PESQUISA. **O Confap** [2019]. Disponível em: <http://confap.org.br/novo/pt/confap>. Acesso em: 07 dez. 2018.

CORDEIRO, D. et al. Da ciência à e-ciência: paradigmas da descoberta do conhecimento. **Revista USP**, São Paulo, n.97, p. 71-80, 2013. Disponível em: <http://www.periodicos.usp.br/revusp/article/view/61867/64710>. Acesso em: 12 dez. 2018.

COSTA, M.; LEITE, F.C.L. Princípios e recomendações basilares para a comunicação dos dados de pesquisa. **Em Questão**, Porto Alegre, v. 23, n. 1, p. 87-112, 2017. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/65623>. Acesso em: 11. Set. 2019.

CHOO, C. W. **Information management for the intelligent organization: the art of scanning the environment**. 3.ed. [s.l.]: ASIST, 2002.

CURTY, R. G. **Beyond "Data Thrifting"**: an investigation of factors influencing research data reuse in the social sciences. 2015. Tese (Doutorado em Filosofia da Ciência da Informação e Tecnologia), Syracuse University. 2015. Disponível em: <https://surface.syr.edu/etd/266/>. Acesso em: 04 nov. 2017.

CURTY, R. G. Factors influencing research data reuse in the social sciences: an exploratory study. **International Journal of Digital Curation**, v. 11, n. 1, p. 96–117, 2016. Disponível em: <http://www.ijdc.net/article/view/11.1.96>. Acesso em: 26 ago. 2018.

CURTY, R. G. et al. Attitudes and norms affecting scientists data reuse. **PLoS ONE**, v.12, n. 12, p. 1-22, 2017. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0189288>. Acesso em: 02 ago. 2018.

CURTY, R. G. Abordagens de reuso e a questão da reusabilidade dos dados científicos. **Liinc em Revista**, v. 15, n. 2, p. 177-193, 2019. Disponível em: doi.org/10.18617/liinc.v15i2.4777. Acesso em: 17 dez. 2020.

DAVID, P.A. **The historical origins of Open Science**: an essay on patronage, reputation and common agency contracting in the scientific revolution. 2000. Disponível em: <http://www-siepr.stanford.edu/workp/swp06008.pdf>. Acesso em: 26 jun. 2019.

DALMORO, M.; VIEIRA, K. M. Dilemas na construção de escalas do tipo Likert: o número de itens e a disposição influenciam nos resultados? **Revista Gestão Organizacional**, v. 6, n. 3, p. 161-174, 2013. Disponível em: <http://www.anpad.org.br/admin/pdf/EPQ-A1615.pdf>. Acesso em: 03 mar. 2019.

DUBLIN CORE METADATA INITIATIVE. **DCMI history**. Disponível em: <https://dublincore.org/about/history/>. Acesso em: 02 fev. 2020.

DEMO, P. Pesquisa social. **Serviço social e Realidade**, Franca, v. 17, n. 1, p. 11-36, 2008. Disponível em: <https://ojs.franca.unesp.br/index.php/SSR/article/view/1>. Acesso em: 18 set. 2019.

DIANA, J. B. **Fatores bióticos e abióticos**. [online], [s/d]. Disponível em: <https://www.todamateria.com.br/fatores-bioticos-e-abioticos/>. Acesso em: 18 set. 2019.

DIGITAL CURATION CENTER. **Research data lifecycle**. Disponível em: <http://www.dcc.ac.uk/>. Acesso em: 01 out. 2018.

EUROPEAN COMMISSION. **Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020**. 2016. Disponível em: <https://goo.gl/RgkJpg>. Acesso em: 07 ago. 2018.

FANIEL; IXCHEL, M.; ZIMMERMAN, A. Beyond the data deluge: a research agenda for large-scale data sharing and reuse. **The International Journal of Digital Curation**, v.6, n. 1, p. 58-69, 2011. Disponível em: <http://www.ijdc.net/article/view/163/231>. Acesso em: 11 ago. 2018.

FREDERICK, D. E. Data, open science and libraries – the data deluge column. **Library Hi Tech News**, n. 8, p. 11-16, 2016. Disponível em: [doi/10.1108/LHTN-09-2016-0040/full/html](https://doi.org/10.1108/LHTN-09-2016-0040/full/html). Acesso em: 04 abr. 2019.

FERRER-SAPENA, A. et al. Cómo analizar el impacto de los datos de investigación on métricas: modelos y servicios. **El Profesional de la Información**, v. 25, n. 4, p. 632-641, 2016. Disponível em: 10.3145/epi.2016.jul.13. Acesso em: 11 mar. 2020.

FISHBEIN, M.; AJZEN, I. Attitudes towards objects as predictors of a single and multiple behavioral criteria. **Psychological Review**, v. 81, n. 1, p. 59-74, 1974. Disponível em: <https://psycnet.apa.org/record/1974-24385-001>. Acesso em: 12 nov. 2019.

FREITAS, H.; JANISSEK-MUNIZ, R.; e MOSCAROLA J. Dinâmica do processo de coleta e análise de dados via web. In: Congresso Internacional de Pesquisa Qualitativa, 1., 2004, Taubaté, SP. **Anais...** Taubaté, SP: CIBRAPEQ, 2004. p. 1-13. Disponível em: http://gianti.ea.ufrgs.br/files/artigos/2004/2004_157_CIBRAPEQ.pdf. Acesso em: 22 fev. 2020.

FREITAS, J.L; BUFREM, L.S.; GRÁCIO, M.C.C. O interdomínio dos estudos métricos da informação em medicina: aproximação entre discurso e prática de seus pesquisadores. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 24, n. 56, 2017. p. 1-22. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2019.e65348/40795>. Acesso em: 12 nov. 2019.

FAPESP. **FAPESP lança rede de repositório de dados científicos do estado de São Paulo**. [20 de dezembro de 2019]. Disponível em: <http://agencia.fapesp.br/fapesp-lanca-rede-de-repositorios-de-dados-cientificos-do-estado-de-sao-paulo/32251/>. Acesso em: 02 jan. 2020.

GIL. A. C. **Como elaborar projetos de pesquisa**. 7.ed. São Paulo: Atlas, 2019.

GIL. A. C. **Métodos e técnicas de pesquisa social**. 6.ed. São Paulo: Atlas, 2008.

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T. Estudos métricos da informação no brasil: análise da interlocução entre os pesquisadores de destaque. **Encontro Brasileiro de Bibliometria e Cientometria**, v. 5, 2016. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/46142>. Acesso em: 23 nov. 2019.

GRIGG, K. S. Data in the Sciences. In: KELLAM, L.; THOMPSON, K. **Datalibrarianship: The Academic Data Librarian in Theory and Practice**. 2015, p. 179-192. Disponível em: https://libres.uncg.edu/ir/uncg/f/K_Grigg_Data_2015.pdf. Acesso em: 08 ago. 2018.

GRINYER, A. The ethics of the secondary analysis and further use of qualitative data. **Social Research Update**, v.56, n.4, 2009. Disponível em: <http://sru.soc.surrey.ac.uk/SRU56.pdf>. Acesso em: 10 set. 2018.

HAIR, J.F.P.J. et al. **Análise multivariada de dados**. São Paulo: Bookman, 2009.

HEY, T.; TANSLEY, S.; TOLLE, K., (edit.) **The fourth paradigm: data-intensive scientific discovery**. Washington, EUA: Microsoft Research, 2009. Disponível em:

<https://www.immagic.com/eLibrary/ARCHIVES/EBOOKS/M091000H.pdf>. Acesso em: 19 ago. 2018.

HIGGINS, S. THE DCC curation lifecycle model. **The International Journal of Digital Curation**, v. 3, n. 1, p. 134-140, 2008. Disponível em: <http://www.ijdc.net/article/view/69>. Acesso em: 11 set. 2019.

HOFELICH MOHR, A.; JOHNSTON, L. R.; LINDSAY, T. A. The data management village: collaboration among research support providers in the large academic environment. In: KELLNA, L.; THOMPSON, K. **Databrarianship: the academic data librarian in theory and practice**. Chicago, IL, EUA: Association of College and Research Libraries, 2016. Disponível em: <http://hdl.handle.net/11299/181127>. Acesso em: 28 jul. 2019.

IBICT. **Manifesto de acesso aberto a dados da pesquisa brasileira para ciência cidadã**. 2016. Disponível em: <http://www.ibict.br/sala-de-imprensa/noticias/item/478-ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada>. Acesso em: 02abr. 2018.

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (ICPSR). **Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle**. 5.ed. Ann Arbor, MI, EUA: [s.n.], 2012.

JONES, M. B et al. **Ecological Metadata Language version 2.2.0**. Disponível em: <https://eml.ecoinformatics.org>. Acesso em: 19 set. 2019.

JORGE, V. A. **Abertura e compartilhamento de dados para pesquisa nas situações de emergência em saúde pública: o caso do vírus Zika**. 2018. 263 f. Tese (Doutorado em Ciência da Informação), Escola de Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018. Disponível em: http://repositorio.ibict.br/bitstream/123456789/993/1/Tese_VanessaJorge.pdf. Acesso em: 13 set. 2019.

KINDLING, M. et al. The landscape of research data repositories in 2015: a re3dta analysis. **D-Lib Magazine**, v. 23, n. ¾, 2017. Disponível em: <http://www.dlib.org/dlib/march17/kindling/03kindling.html>. Acesso em: 09 jun. 2019.

KUHN, T.S. **A estrutura das revoluções científicas**. 5.ed. São Paulo: Editora Perspectiva, 1982.

LATOUR, B. **Ciência em ação: como seguir cientistas e engenheiros sociedade a fora**. São Paulo: UNESP, 2000, p. 201-246.

LATOUR, B. Um coletivo de humanos e não-humanos: no labirinto de Dédalo. In: LATOUR, B. **A esperança de Pandora: ensaio sobre a realidade dos estudos científicos**. Bauru, SP: EDSUP, 2001.

LIMA, V.M.L.M.; D'AMORIM, M.A.M. A relação atitude-comportamento à luz da Teoria da Ação Racional. **Arquivos Brasileiros de Psicologia**, v. 38, n. 1, p. 133-

142, 1986. Disponível em:
<http://bibliotecadigital.fgv.br/ojs/index.php/abp/article/viewFile/19238/17972>. Acesso em: 12 nov. 2019.

LYNCH, C. Jim Gray's fourth paradigm and the construction of the scientific record. IN: **The fourth paradigm: data-intensive scientific discovery**. Washington, EUA: Microsoft Research, p. 177-183, 2009. Disponível em:
<https://www.immagic.com/eLibrary/ARCHIVES/EBOOKS/M091000H.pdf>. Acesso em: 19 dez. 2018.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 8.ed. São Paulo, Atlas, 2017.

MEDEIROS, J. da S.; CAREGNATO, Sônia Elisa. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, Rio de Janeiro, v.8, n. 2, p. 311-322, 2012. Disponível em:
<http://revista.ibict.br/liinc/article/view/3368>. Acesso em: 19 set. 2019.

MERTON, R. K. **The sociology of science**: theoretical and empirical investigations. Chicago: University of Chicago Press, 1973.

MOUTINHO, K.; ROASSI, A. As teorias da ação racional e teoria da ação planejada: relações entre intenções e comportamentos. **Avaliação Psicológica**, v. 9, n. 2, 2010. p. 279-287. Disponível em:
http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000200012. Acesso em: 12 nov. 2019.

MURRAY-RUST, P. et al. **Panton Principles, Principles for open data in science**. 2010. Disponível em: <https://pantonprinciples.org/>. Acesso em: 10 mar. 2019.

NATIONAL SCIENCE FOUNDATION. **Cyberinfrastructure Vision for 21st Century Discovery**. 2007. Disponível em: <https://www.nsf.gov/pubs/2007/nsf0728/>. Acesso em: 08 ago. 2018.

NIELSEN, H.J.; HJØRLAND, B. Curating research data: the potential of roles of libraries and information professionals. **Journal of Documentation**, v. 70, n. 2, 2014. p. 221-240. Disponível em: 10.1108/JD-03-20130034/full/html#idm45364477961952. Acesso em: 17 mar. 2020.

OLIVEIRA, A. C. S. Adoção de padrões de metadados para repositórios de dados digitais na ciência aberta. In: VECHIATTO, A. et al. (orgs.). **Repositórios digitais: teoria e prática**. Curitiba: EDUTFPR, 2017. p. 167-192. Disponível em: <http://repositorio.utfpr.edu.br/jspui/>. Acesso em: 19 out. 2019.

OLIVEIRA, Adriana Carla Silva de; SILVA, E. M. da. Ciência aberta: dimensões para um novo fazer científico. **Informação & Informação**, Londrina, v. 21, n. 2, p. 5-39, dez. 2016. Disponível em:
<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27666/20113>. Acesso em: 19 ago. 2018.

OPEN KNOWLEDGE FOUNDATION. **Is it open data?** [2009]. Disponível em: <http://www.isitopendata.org/>. Acesso em: 6 mar. 2019.

PAGANINE, L.; COSTA, M. Repositório de dados de pesquisa para as ciências da saúde. **Cadernos BAD**, n. 2, 2016, p. 57-70. Disponível em: <https://www.bad.pt/publicacoes/index.php/cadernos/article/view/1583/pdf>. Acesso em: 01 out. 2019.

PAVÃO, C. G.; ROCHA, R. P.; GABRIEL JUNIOR, R. F. Proposta de criação de uma rede de dados abertos da pesquisa brasileira. **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 16, n. 2, p. 329-343, 2018. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8651180>. Acesso em: 10 jun. 2019.

PIWOWAR, H. A.; DAY, R. S.; FRIDSMA, D. B. Sharing detailed research data is associated with increased citation rate. **PloS ONE**, v. 2, n. 3, 2007. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308>. Acesso em: 20 abr. 2019.

PIWOWAR, H. A.; CHAPMAN, W. W. A review of journal policies for sharing research data. In: ELPUB 2008 CONFERENCE ON ELETRONIC PUBLISHING, 2008. Toronto, Canadá. **Proceedings...** Toronto, Canadá. 2008. Disponível em: https://elpub.architexturez.net/system/files/pdf/001_elpub2008.content.pdf. Acesso em: 21 abr. 2019.

ORGANIZAÇÃO PARA COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. **Making Open Science a Reality**. (OECD Science, Technology and Industry Policy Papers, 25). Paris: OECD Publishing. 2015. Disponível em: <http://dx.doi.org/10.1787/5jrs2f963zs1-en>. Acesso em: 09 ago. 2018.

RE3DATA. Disponível em: <https://www.re3data.org>. Acesso em: 27 out. 2018.

RODRIGUES et al. **Os repositórios de dados científicos: o estado da arte**. 2010. Disponível em: <http://repositorium.sdum.uminho.pt/handle/1822/10830>. Acesso em: 01 out. 2019.

RÜCKNAGEL et al. **Schema for the description of research data repositories: version 3.0**. 2015. Disponível em: <http://doi.org/10.2312/re3.004>. Acesso em: 02 out. 2019.

SALES, L. F. **Integração semântica de publicações científicas e de dados de pesquisa**: proposta de modelo de publicação ampliada para a área de ciências nucleares. Tese (Doutorado em Ciência da Informação), Programa de Pós-graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Disponível em: <https://ridi.ibict.br/bitstream/123456789/874/1/LUANA%20SALES%20D.pdf>. Acesso em: 30 nov. 2019.

SANCHEZ, F. A.; VECHIATO, F.L. Avaliação de repositórios de dados da pesquisa segundo critérios da encontrabilidade da informação. **Liinc em Revista** v. 15, n. 2, p. 142-160, 2019. Disponível em: <https://doi.org/10.18617/liinc.v15i2.4774>. Acesso em: 07 jan. 2020.

SANTOS, P. X. (Coord.). **Livro Verde - Ciência aberta e dados abertos**: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional. Rio de Janeiro: Fiocruz, 2017. 141 p. Disponível em: <https://www.arca.fiocruz.br/handle/icict/24117>. Acesso em: 10 mar. 2019.

SAYÃO, L.F. Outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, v. 15, n. 30, p. 1-31, 2010. Disponível em: 10.5007/1518-2924.2010v15n30p1. Acesso em: 08 dez. 2018.

SAYÃO, L.F.; SALES, L. F. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v.17, nesp. 2, III SBCC, p. 118-135, 2012. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp2p118>. Acesso: 08 dez. 2018.

SAYÃO, L.F.; SALES, L. F. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, Belo Horizonte, v. 6, n.1, 2013. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/pbcib/article/view/18634>. Acesso em: 13 set. 2019.

SAYÃO, L.F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. 90 p. Disponível em: http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 10 fev. 2018.

SAYÃO, L.F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Inf. Inf.**, Londrina, v.21, n.2, p. 90-115, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 08 dez. 2018.

SILVA, F. C. C. da. **Gestão de dados científicos**. Rio de Janeiro: Editora Interciência, 2019.

SMITH, G. D. Increasing the accessibility of data. **The BMJ**, v. 308, 11 jun., 1994, p. 1519-1520. Disponível em: <https://www.bmj.com/content/308/6943/1519.full>. Acesso: 14 set. 2019.

SMITH, P. Data science as an academic discipline. **Data Science Journal**, n.5, p. 163-164. Disponível em: <https://datascience.codata.org/articles/abstract/10.2481/dsj.5.163/>. Acesso em: 11 dez. 2018.

SUTTON, J.; AUSTIN, Z. Qualitative research: data collection, analysis, and management. **The Canadian Journal of Hospital Pharmacy**, v. 68, n. 3, 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4485510/>. Acesso em: 10 mar. 2019.

TENOPIR, C. et al. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. **PLoS One**, v.10, n.8, 2015. Disponível em: <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0134826&type=printable>. Acesso em: 17 ago. 2018.

TENOPIR, C. et al. Data sharing by scientists: practices and perceptions. **PLoS ONE**, v. 6, n.6, 2011. Disponível em: doi:10.1371/journal.pone.0021101. Acesso em: 17 set. 2018.

TOMAÉL, M. I.; MARTELETO, R. M. Redes sociais: posições dos autores no fluxo da informação. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 11, n. 1, nesp, p. 75-91, 2006. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2006v11nesp1p75/387>. Acesso em: 08 dez. 2018.

VANZ, S. A. S. et al. **Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores: relatório 2018**. Disponível em: <http://hdl.handle.net/10183/185195>. Acesso em: 01 out. 2019.

VILLAVECES, J.L. et al. ¿Cómo medir el impacto de las políticas de ciencia y tecnología? **Revista CTS**, v. 4, n. 2, p. 125-146, 2005. Disponível em: <http://www.revistacts.net/files/Volumen%202%20-%20N%FAmero%204/doss04.pdf>. Acesso em: 07 jan. 2020.

WILKINSON, M.D., et al. **The FAIR Guiding Principles for scientific data management and stewardship**. *Scientific Data*, v.3, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 17 set. 2019.

WU, M. et al. Data discovery paradigms: user requirements and recommendations for data repositories. **Data Science Journal**, v. 18, n. 3, p. 1-13, 2019. Disponível em: <https://datascience.codata.org/articles/10.5334/dsj-2019-003/>. Acesso em: 30 abr. 2019.

APÊNDICE 1 – QUESTIONÁRIO

IDENTIFICAÇÃO DOS FATORES DE IMPACTO PERCEBIDO PELOS PESQUISADORES QUANTO AO REUSO DE DADOS CIENTÍFICOS EM REPOSITÓRIOS BRASILEIROS

Carta de apresentação para coleta de dados
Curitiba, 01 de janeiro de 2020.

Prezado(a) Pesquisador(a):

Sou mestranda do Programa de Pós-Graduação em Gestão da Informação pela Universidade Federal do Paraná (UFPR) e estou realizando uma pesquisa de campo para minha dissertação com o objetivo de identificar o impacto percebido pelos pesquisadores sobre o reuso de dados científicos nos repositórios brasileiros registrados no Re3data.org.

Para o alcance dos objetivos propostos necessito da colaboração dos pesquisadores que contém dados ou conjuntos de dados científicos depositados em repositório de dados nacionais.

O questionário possui fins exclusivamente acadêmicos e tem o objetivo de identificar o impacto percebido pelos pesquisadores quanto ao reuso de dados científicos de pesquisa. Todas as informações referentes a identificação dos respondentes serão sigilosas e descaracterizadas. Os dados coletados serão publicados no relatório de pesquisa e disponibilizados de forma aberta em repositório de dados institucional da UFPR. Os respondentes não serão identificados.

Solicitamos a gentileza para que o questionário seja preenchido no link abaixo até o dia 10 de fevereiro de 2020.

Qualquer dúvida ou questionamento, por favor entrar em contato pelo e-mail.

Agradeço pela atenção e disponibilidade.

**Termo de
Consentimento
Livre e
Esclarecido
(TCLE)**

Você está sendo convidado(a) a participar da pesquisa de mestrado “Impacto percebido pelos pesquisadores quanto ao reuso de dados científicos de pesquisa em repositórios de dados brasileiros”.

O objetivo desta pesquisa é identificar o impacto percebido pelos pesquisadores sobre o reuso de dados científicos nos repositórios brasileiros registrados no Re3data.

Caso você aceite participar da pesquisa, será necessário responder ao questionário on-line, desenvolvido pela autora desta pesquisa.

O tempo necessário para responder o questionário é de aproximadamente 10 minutos.

A sua participação é voluntária e não envolve custos financeiros. Todas as respostas serão tratadas de forma anônima e você não será identificado(a). Os dados coletados serão despersonalizados e serão publicados na dissertação e também disponibilizados de forma aberta em repositório de dados institucional da UFPR sem identificá-lo(a). Estão garantidas todas as informações de seu interesse a qualquer momento da pesquisa. Caso você desista de participar, basta solicitar de volta este termo de consentimento.

Eu, abaixo identificado(a) e assinado, li o texto acima e compreendi a natureza e objetivo do estudo do qual fui convidado a participar. A explicação que recebi menciona todos os aspectos compreendidos pela pesquisa. Eu entendi que sou livre para interromper minha participação na pesquisa a qualquer momento sem a necessidade de justificar minha decisão. Eu concordo voluntariamente em participar deste estudo. O Termo de Consentimento Livre e Esclarecido, poderá ser enviado eletronicamente após a sua assinatura, ou apenas confirmando seu consentimento por e-mail a pesquisadora.

Nome completo do respondente:

Assinatura:

Parte I - Identificação dos respondentes.

1. 1. Qual é a sua principal ocupação? *

Marcar apenas uma oval.

- ☐ Aluno de graduação
- ☐ Aluno de pós-graduação
- ☐ Docente
- ☐ Pesquisador
- ☐ Bibliotecário
- ☐ Gestor
- ☐ Outro: _____

2. 2. Em qual das seguintes regiões brasileiras está situado seu principal local de trabalho? * *Marcar apenas uma oval.*

- ☐ Centro-oeste
- ☐ Nordeste
- ☐ Norte
- ☐ Sudeste
- ☐ Sul
- ☐ Outro:

3. 3. Qual é a sua área de atuação? *

Marcar apenas uma oval.

- ☐ Ciências agrárias
- ☐ Ciências biológicas
- ☐ Ciências humanas
- ☐ Ciências da saúde
- ☐ Ciência sociais aplicadas
- ☐ Engenharias
- ☐ Linguística, letras e artes
- ☐ Bioética
- ☐ Ciências ambientais
- ☐ Defesa
- ☐ Divulgação científica

4. 4. O que você diria ser a sua principal orientação/abordagem metodológica para pesquisa?

Marcar apenas uma oval.

- ☐ Qualitativa
- ☐ Quantitativa
- ☐ Mista

5. 5. Qual dos seguintes setores você está vinculado? *

136

Marcar apenas uma oval.

- ☐ Acadêmico público
- ☐ Acadêmico privado
- ☐ Institutos de pesquisa
- ☐ Iniciativa privada
- ☐ Sem vínculo formal
- ☐ Outro: _____

6. 6. Qual das alternativas melhor descreve o principal financiador da sua pesquisa? *

Marcar apenas uma oval.

- ☐ Agência estadual
- ☐ Agência nacional
- ☐ Agência internacional
- ☐ Iniciativa privada
- ☐ A instituição profissional ao qual está vinculado
- ☐ Financiamento particular próprio
- ☐ Outro: _____

7. 7. Qual é o seu nível de escolaridade? *

Marcar apenas uma oval.

- ☐ Graduação
- ☐ Mestrado
- ☐ Doutorado
- ☐ Pós-doutorado
- ☐ Sem formação acadêmica

8. 8. Qual termo melhor descreve o tipo de dado que você usa? *

137

Marque todas que se aplicam.

- ☐ Dados abióticos (solos, microclima, hidrologia, etc.)
- ☐ Dados bióticos
- ☐ Modelos de dados
- ☐ Dados experimentais (envolvendo algum tipo de manipulação)
- ☐ Dados observacionais (sem envolver qualquer tipo de manipulação)
- ☐ Dados abióticos com sensoriamento remoto (incluindo dados meteorológicos)
- ☐ Pesquisas em ciências sociais
- ☐ Entrevistas

Outro: ☐ _____

9. 9. Se alguns ou todos os seus dados estão disponíveis para outras pessoas, onde eles estão disponíveis?

Marcar apenas uma oval.

- ☐ No site da instituição ao qual estou vinculado
- ☐ No site do líder do meu grupo de pesquisas
- ☐ Através de uma rede regional
- ☐ Através de uma rede nacional
- ☐ Repositório de dados da minha instituição
- ☐ No meu site pessoal
- ☐ Outro: _____

17. Outros podem acessar facilmente meus dados. *

140

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

18. Estou satisfeito com as ferramentas para preparar meus dados. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

19. Estou satisfeito com as ferramentas para preparar a documentação dos meus dados. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

12. Como a instituição que você é vinculado/trabalha está envolvida com seus dados?

Diga o quanto você discorda em cada afirmação utilizando a seguinte escala: 1. Discordo fortemente, 2. Discordo, 3. Nem discordo nem concordo, 4. Concordo, 5. Concordo fortemente.

20. Minha instituição ou grupo de pesquisa tem um processo formal estabelecido para gestão de dados durante a vida do projeto (curto prazo).

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

21. Minha instituição ou grupo de pesquisa tem um processo formal estabelecido para gestão de dados para além da vida do projeto (longo prazo).

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

22. Minha instituição ou grupo de pesquisa tem um processo formal estabelecido para armazenar os dados para além da vida do projeto (longo prazo).

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

23. Minha instituição ou grupo de pesquisa provê as ferramentas e suporte técnico para a gestão de dados durante a vida da pesquisa (curto prazo).

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

24. Minha instituição ou grupo de pesquisa provê as ferramentas e suporte técnico para a gestão de dados para além da vida da pesquisa (longo prazo).

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

25. Minha instituição ou grupo de pesquisa provê treinamentos e melhores práticas em gestão de dados.

142

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

26. Minha instituição ou grupo de pesquisa provê recursos necessários para a gestão de dados durante a vida do projeto (curto prazo).

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

27. Minha instituição ou grupo de pesquisa provê recursos necessários para a gestão de dados para além da vida do projeto (longo prazo).

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

13. As questões a seguir tratam sobre as suas opiniões sobre uso de dados em sua área de pesquisa.

Diga o quanto você discorda em cada afirmação utilizando a seguinte escala: 1. Discordo fortemente, 2. Discordo, 3. Nem discordo nem concordo, 4. Concordo, 5. Concordo fortemente.

28. A falta de acesso aos dados gerados por outros pesquisadores ou instituições é o princip impedimento para o progresso da ciência.

143

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

29. A falta de acesso aos dados gerados por outros pesquisadores ou instituições restringe a minha habilidade de responder a questionamentos.

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

30. Os dados podem ser mal interpretados devido a sua complexidade.

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

31. Os dados podem ser mal interpretados devido a sua baixa qualidade.

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

32. Os dados podem ser usados de maneiras adversas além das pretendidas.

144

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

Parte III - Comportamento do pesquisador quanto ao reuso de dados científicos de pesquisa.

14. As questões a seguir referem-se ao comportamento do pesquisador quanto ao reuso de dados.

Diga o quanto você discorda em cada afirmação utilizando a seguinte escala: 1. Discordo fortemente, 2. Discordo, 3. Nem discordo nem concordo, 4. Concordo, 5. Concordo fortemente.

33. Eu usaria conjuntos de dados de outros pesquisadores se eles fossem facilmente acessíveis.

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

34. Eu estaria disposto a colocar pelo menos alguns dos meus dados em um repositório de dados sem restrições.

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

35. Eu estaria disposto a colocar todo os meus dados em um repositório de dados sem restrições.

145

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

36. Estaria mais favorável a colocar meus dados disponíveis se eu pudesse acrescentar condições de acesso.

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

37. Estou satisfeito com minha habilidade em integrar dados de diferentes fontes para minha questões de pesquisa.

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

38. Estaria disposto a compartilhar dados em um amplo grupo de pesquisadores que utilizam dados de diferentes formas.

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

39. É importante que meus dados sejam citados quando utilizados por outros pesquisadores *Marcar apenas uma oval.*

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

40. É apropriado criar novos conjuntos de dados a partir de dados compartilhados.
Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

41. 15. Que padrão de metadados você utiliza para descrever seus dados? *

Marque todas que se aplicam.

- ☐ Dubin Core (DC)
- ☐ Darwin Core (DwC)
- ☐ Directory Interchange Format (DIF)
- ☐ Ecological Metadata Language (EML)
- ☐ International Standards Organization (ISO)
- ☐ OpenGIS (OGIS)
- ☐ Padrão de metadados próprio desenvolvido no meu laboratório ou pelo meu grupo de pesquisa
- ☐ Não sei o que é um metadado
- ☐ Nenhum

Outro: ☐ _____

42. 16. Você é o único responsável por aprovar acesso aos seus dados? *

Marcar apenas uma oval.

- ☐ Sim, para todos os meus conjuntos de dados
- ☐ Sim, para alguns dos meus conjuntos de dados
- ☐ Não

43. 17. A principal agência de financiamento da sua pesquisa exige o fornecimento de um Plan de Gestão de Dados (PGD)? *

Marcar apenas uma oval.

- ☐ Sim
- ☐ Não
- ☐ Não sei

44. 18. Na sua opinião, o reuso de dados... *

Marcar apenas uma oval.

- ☐ Permite a verificação de resultados de pesquisas anteriores
- ☐ Contribui para o avanço científico
- ☐ Permite a reprodução de pesquisas
- ☐ Reduz os custos associados com a pesquisa
- ☐ Poupa o tempo do reutilizador

19. Na sua opinião, ao usar dados de outras pessoas em sua própria pesquisa, você poderia correr o risco de...

Diga o quanto você discorda em cada afirmação utilizando a seguinte escala: 1. Discordo fortemente, 2. Discordo, 3. Nem discordo nem concordo, 4. Concordo, 5. Concordo fortemente.

45. Receber reconhecimento inferior. *

Marcar apenas uma oval.

1 2 3 4 5

Discordo fortemente ☐ ☐ ☐ ☐ ☐ Concordo fortemente

46. Ter minha pesquisa mal considerada. *

148

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

47. Infringir involuntariamente os códigos éticos. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

48. Violar involuntariamente as normas de proteção de dados. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

49. Não ter consentimento informado pelos participantes de uma pesquisa. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

50. Interpretar mal os dados. *

149

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

51. Fazer suposições ou associações incorretas com base nos dados. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

52. Usar dados indevidos. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

53. Ficar vulnerável a dados com erros ocultos. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

54. Ter capacidade limitada para identificar erros nos dados. *

150

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

55. Reutilizar dados imprecisos. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

20. Na sua opinião, o reuso de dados requer muito esforço dos pesquisadores para...

Diga o quanto você discorda em cada afirmação utilizando a seguinte escala: 1. Discordo fortemente, 2. Discordo, 3. Nem discordo nem concordo, 4. Concordo, 5. Concordo fortemente.

56. Encontrar formas inovadores de abordar dados existentes. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

57. Justificar a importância de um novo estudo com base nos dados existentes. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

58. Identificar novos dados relevantes. *

151

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

59. Selecionar dados reutilizáveis. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

60. Obter permissão para reutilizar dados. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

61. Recuperar dados e familiarizar-se com eles. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

62. Coletar e produzir dados por outras pessoas. *

152

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

63. Compreender o estudo original. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

64. Ajustar seu próprio projeto de pesquisa para acomodar os dados existentes. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

65. Alterar uma ideia de pesquisa para acomodar os dados existentes *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

66. Reorganizar seu estudo inicial para acomodar os dados existentes. *

153

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

67. 21. Quais oportunidades você percebe no reuso de dados científicos de pesquisa? *

Marque todas que se aplicam.

- ☐ a possibilidade de reprodução ou verificação de resultados
- ☐ tornar disponível de forma aberta o resultado de pesquisas realizadas com recursos públicos
- ☐ permitir que outras pessoas façam novas perguntas baseadas nos dados existentes
- ☐ avançar no estado de pesquisa e inovação
- ☐ benefícios na carreira (por exemplo: reconhecimento e boa reputação)
- ☐ aumento do número de citações
- ☐ convites para colaboração com outros pesquisadores ou grupos de pesquisa
- ☐ a possibilidade de transformação social na ciência cidadã

Outro: ☐ _____

22. Na sua opinião, para serem reutilizáveis, os dados de pesquisa existentes devem...

Diga o quanto você discorda em cada afirmação utilizando a seguinte escala: 1. Discordo fortemente, 2. Discordo, 3. Nem discordo nem concordo, 4. Concordo, 5. Concordo fortemente.

68. Ter documentação suficiente (por exemplo: livro de códigos). *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

69. Ter documentação simples e clara. *

154

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

70. Estar na mesma unidade de análise que eu preciso (por exemplo: indivíduos, grupos). *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

71. Encaixar no meu estudo. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

72. Estar conforme o formato que eu preciso. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

77. Ser o produto de um estudo bem projetado. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

78. Pertencer a um estudo que realizou um plano de pesquisa original. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

79. Ser originado de um estudo bem executável. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

23. Sobre reuso de dados, eu diria que...

Diga o quanto você discorda em cada afirmação utilizando a seguinte escala: 1. Discordo fortemente, 2. Discordo, 3. Nem discordo nem concordo, 4. Concordo, 5. Concordo fortemente.

80. Estou ciente dos repositórios de onde posso reutilizar dados. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

85. Tenho acesso a treinamento formal sobre habilidades úteis para reutilizar dados. * *Marcar apenas uma oval.*

158

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

24. Na sua opinião, a sua área de pesquisa...

Diga o quanto você discorda em cada afirmação utilizando a seguinte escala: 1. Discordo fortemente, 2. Discordo, 3. Nem discordo nem concordo, 4. Concordo, 5. Concordo fortemente.

86. É receptiva quando ao reuso de dados. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

87. Tem o reuso de dados como prática comum. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

88. Convencionalmente reusa dados. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

89. Não tem interesse em reusar dados. *

159

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

25. Eu poderia...

Diga o quanto você discorda em cada afirmação utilizando a seguinte escala: 1. Discordo fortemente, 2. Discordo, 3. Nem discordo nem concordo, 4. Concordo, 5. Concordo fortemente.

90. Incentivar-me a reusar dados. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

91. Apoiar o reuso de dados. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

92. Reusar dados frequentemente. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

APÊNDICE 2 – POLÍTICAS E/OU DOCUMENTOS NORTEADORES SOBRE GESTÃO DE DADOS CIENTÍFICOS PELAS FUNDAÇÕES ESTADUAIS DE AMPARO À PESQUISA

TABELA 16 - POLÍTICAS E/OU DOCUMENTO NORTEADORES SOBRE GESTÃO DE DADOS CIENTÍFICOS PELAS FUNDAÇÕES DE AMPARO À PESQUISA ESTADUAIS

Entidade	Política	Indicação de Repositório
CNPq	Chamada no 38/2018 CNPQ/EQUINOR ENERGIA LTDA. 2018 - Formação de Mestres e Doutores em Petróleo, Gás Natural e Energia Renovável. (2018)	Não há
	Chamada no 39/2018 CNPq/Instituto Aggeu Magalhães-FIOCRUZ - Apoiar atividades de pesquisa científica, tecnológica e de inovação na área de saúde, que visem contribuir significativamente para o desenvolvimento científico, tecnológico e a inovação do País, por meio do Programa de Excelência em Pesquisa – PROEP/IAM. (2018)	Não há
	Chamada no 40/2018 MCTIC/CNPq – Apoio à produção de inventários de Avaliação de Ciclo de Vida. (2018)	Não há
	Chamada no 41/2018 CNPq/Gerência Regional de Brasília – FIOCRUZ - Apoiar projetos de pesquisa científica, tecnológica e de inovação na área de saúde, que visem contribuir significativamente para o desenvolvimento científico, tecnológico e a inovação do País, por meio do Programa de Excelência em Pesquisa em Saúde da Fiocruz Brasília. (2018)	Não há
	Chamada CNPq no 1/2019 - Apoio à formação de doutores em áreas estratégicas. (2018)	Não há
EMBRAPA	Política de Governança de Dados, Informação e Conhecimento da EMBRAPA (2019)	Não há
FAPEMIG	Chamada FAPEMIG 10/2018 – Pesquisa, Desenvolvimento e Inovação em Ambientes Aquáticos de Minas Gerais em Áreas Impactadas pelo Rompimento da Barragem de Fundão, Mariana – MG (2018)	Repositório do SIBBR

FAPESP	Código de Boas Práticas Científicas (2014)	Não há
	Plano de Gestão de Dados (2017)	Modelo FAPESP no DPMT _{ool}
	Política para Acesso Aberto às Publicações Resultantes de Auxílios e Bolsas FAPESP (2019)	Repositórios da CRUESP ou revistas de acesso aberto do SHERPA/RoMEO
FIOCRUZ	Portaria de Acesso Aberto ao Conhecimento Científico da Fiocruz – Portaria da Presidência nº329/2014 (2014)	Repositório institucional ARCA
	Gestão e Abertura de Dados para Pesquisa na Fiocruz – Termo de Referência (2018)	Não há

FONTE: A autora (2019) com informações dos sítios eletrônicos das fundações.

APÊNDICE 3 – CÁLCULO DO COEFICIENTE DO ALFA DE CRONBACH

	Média de escala se o item for excluído	Variância de escala se o item for excluído	Correlação de item total corrigida	Alfa de Cronbach se o item for excluído
Qual é a sua principal ocupação?	296,21	341,649	,043	,713
Qual é a sua região de trabalho?	295,96	345,201	-,059	,720
Qual é a sua área de atuação?	294,94	325,823	,124	,716
Principal tipo de abordagem metodológica?	296,99	341,000	,080	,711
Qual setor profissional está vinculado?	297,84	341,439	,051	,712
Qual é o principal financiador da sua pesquisa?	295,93	330,531	,157	,710
Qual é o tipo de dado que você mais utiliza?	295,31	329,871	,093	,717
Por que todos os seus dados não estão disponíveis?	291,43	315,350	,296	,701
Estou satisfeito com o meu processo de coleta de dados.	294,51	340,543	,108	,710
Estou satisfeito com o processo para coletar os meus dados.	294,80	337,901	,195	,708
Estou satisfeito com o processo para armazenar meus dados durante a vida do projeto (curto prazo).	294,69	336,103	,216	,707
Estou satisfeito com o processo para armazenar meus dados durante a vida do projeto (curto prazo).	294,71	327,164	,389	,700
Estou satisfeito com o processo para análise dos meus dados.	294,49	337,152	,255	,707
Estou satisfeito com o processo para análise dos meus dados.	294,73	322,548	,520	,696

Outros podem acessar facilmente meus dados.	295,11	322,885	,532	,696
Estou satisfeito com as ferramentas para preparar meus dados.	294,89	328,219	,469	,700
Estou satisfeito com as ferramentas para preparar a documentação dos meus dados.	294,99	333,116	,302	,705
Minha instituição ou grupo de pesquisa tem um processo formal estabelecido para gestão de dados durante a vida do projeto (curto prazo).	295,30	332,677	,251	,705
Minha instituição ou grupo de pesquisa tem um processo formal estabelecido para gestão de dados durante a vida do projeto (curto prazo).	295,59	328,275	,364	,701
Minha instituição ou grupo de pesquisa tem um processo formal estabelecido para armazenar os dados para além da vida do projeto (longo prazo).	294,54	318,802	,463	,695
Minha instituição ou grupo de pesquisa provê as ferramentas e suporte técnico para gestão de dados durante a vida do projeto (curto prazo).	295,00	329,130	,344	,702
Minha instituição ou grupo de pesquisa provê as ferramentas e suporte técnico para gestão de dados para além da vida da pesquisa (longo prazo).	295,39	325,893	,416	,699
Minha instituição ou grupo de pesquisa provê treinamentos e melhores práticas em gestão de dados.	295,86	327,950	,395	,701

Minha instituição ou grupo de pesquisa provê necessários para a gestão de dados durante a vida do projeto (curto prazo).	295,56	331,960	,304	,704
Minha instituição ou grupo de pesquisa provê recursos necessários para a gestão de dados para além da vida do projeto (longo prazo).	295,57	326,683	,442	,699
A falta de acesso aos dados gerados por outros pesquisadores ou instituições é o principal impedimento para o progresso da ciência.	295,17	346,376	-,084	,717
A falta de acesso aos dados gerados por outros pesquisadores ou instituições restringe a minha habilidade de responder a questionamentos.	294,90	346,729	-,095	,717
Os dados podem ser mal interpretados devido a sua complexidade.	294,99	344,246	-,025	,714
Os dados podem ser mal interpretados devido a sua baixa qualidade.	294,63	345,280	-,056	,715
Os dados podem ser usados de maneiras adversas além das pretendidas.	294,99	342,739	,026	,713
Eu usaria conjuntos de outros pesquisadores se eles fossem facilmente acessíveis.	293,74	339,092	,236	,708
Eu estaria disposto a colocar pelo menos alguns dos meus dados em um repositório de dados sem restrições.	293,90	332,758	,314	,704
Eu estaria disposto a colocar todos os meus dados em um repositório de dados sem restrições.	294,39	327,748	,335	,702

Estaria mais favorável a colocar meus dados disponíveis se eu pudesse acrescentar condições de acesso.	294,19	351,835	-,201	,723
Estou satisfeito com minha habilidade em integrar dados de diferentes fontes para minhas questões de pesquisa.	294,66	331,214	,407	,702
Estaria disposto a compartilhar dados em um amplo grupo de pesquisadores que utilizam dados de diferentes formas.	293,84	334,424	,443	,704
É importante que meus dados sejam citados quando utilizados por outros pesquisadores.	293,46	344,687	-,041	,713
É apropriado criar novos conjuntos de dados a partir de dados compartilhados.	293,73	337,186	,289	,707
Qual é o principal padrão de metadados utilizado para descrever seus dados?	296,34	346,287	-,092	,734
Você é o único responsável por aprovar acesso aos seus dados?	297,14	344,646	-,036	,714
A principal agência de financiamento da sua pesquisa exige o fornecimento de um Plano de Gestão de Dados (PGD)?	297,91	346,630	-,103	,716
Permite a verificação de resultados de pesquisa anteriores.	297,57	338,799	,322	,708
Contribui para o avanço científico.	297,39	344,675	-,054	,713
Permite a reprodução de pesquisas.	297,57	338,075	,368	,707
Reduz os cursos associados com a pesquisa.	297,64	343,566	,024	,712

Poupa o tempo do reutilizador.	297,87	337,447	,355	,707
Receber reconhecimento inferior.	296,87	339,824	,181	,709
Ter minha esquisa mal considerada.	296,86	338,501	,212	,708
Infringir involuntariamente os códigos éticos.	296,39	334,617	,271	,706
Violar involuntariamente as normas de proteção de dados.	296,27	333,795	,278	,705
Não ter consentimento informado pelos participantes de uma pesquisa.	296,01	339,522	,098	,711
Interpretar mal os dados.	295,50	338,514	,131	,710
Fazer suposições ou associações incorretas com base nos dados.	295,34	339,707	,106	,711
Usar dados indevidos.	295,87	339,621	,109	,710
Ficar vulnerável a dados com erros ocultos.	294,81	340,994	,083	,711
Ter capacidade limitada para identificar erros nos dados.	294,80	339,641	,120	,710
Reutilizar dados imprecisos.	294,94	336,229	,195	,708
Encontrar formas inovadores de abordar dados existentes.	295,06	341,649	,040	,713
Justificar a importância de um novo estudo com base nos dados existentes.	295,20	344,075	-,024	,715
Identificar novos dados relevantes.	294,17	348,144	-,123	,719
Selecionar dados reutilizáveis.	293,71	342,758	,029	,712
Obter permissão para reutilizar dados.	294,91	347,993	-,145	,717
Recuperar dados e familiarizar-se com eles.	294,79	347,330	-,126	,716
Coletar e produzir dados por outras pessoas.	294,84	349,062	-,176	,718
Compreender o estudo original.	295,29	345,251	-,055	,715

Ajustar seu próprio projeto de pesquisa para acomodar os dados existentes.	294,81	340,849	,093	,711
Alterar uma ideia de pesquisa para acomodar os dados existentes.	294,86	339,081	,150	,709
Reorganizar seu estudo inicial para acomodar os dados existentes.	294,87	341,621	,058	,712
a possibilidade de reprodução ou verificação de resultados.	297,63	339,454	,264	,708
tornar disponível de forma aberta o resultado de pesquisas realizadas com recursos públicos	297,41	345,377	-,111	,713
permitir que outras pessoas façam novas perguntas baseadas nos dados existentes	297,39	342,153	,208	,710
avançar no estado de pesquisa e inovação	297,57	339,611	,272	,708
benefícios na carreira (por exemplo: reconhecimento e boa reputação)	297,84	338,482	,296	,708
aumento do número de citações	297,71	336,990	,386	,706
convites para colaboração com outros pesquisadores ou grupos de pesquisa	297,77	340,933	,164	,710
a possibilidade de transformação social na ciência cidadã	297,90	343,454	,028	,712
Ter documentação suficiente (por exemplo: livro de códigos).	293,76	343,926	-,008	,713
Ter documentação simples e clara.	293,57	345,611	-,097	,714
Estar na mesma unidade de análise que eu preciso (por exemplo: indivíduos, grupos).	294,13	339,708	,094	,711
Encaixar no meu estudo.	294,01	333,493	,272	,705

Estar conforme o formato que eu preciso.	294,46	341,005	,051	,713
Ser originário de uma fonte confiável.	293,64	344,697	-,039	,713
Ser produto de uma fonte confiável.	294,19	346,095	-,082	,716
Ser consistente/estável.	294,09	340,398	,104	,711
Estar completo ou com poucos dados ausentes.	293,60	341,548	,096	,711
Ser produto de um estudo bem projetado	294,71	343,888	-,013	,714
Pertencer a um estudo que realizou um plano de pesquisa original.	295,60	335,200	,196	,707
Ser originado de um estudo bem executável.	294,87	344,461	-,031	,714
Estou ciente dos repositórios de onde posso reutilizar dados.	294,81	336,936	,197	,708
Posso encontrar facilmente repositórios com dados relacionados à minha pesquisa.	295,21	337,446	,208	,708
Os pesquisadores primários estão disponíveis para fornecer informações adicionais aos seus dados.	296,01	335,580	,235	,707
Posso facilmente entrar em contato com os pesquisadores primários para obter esclarecimentos sobre seus dados, se eu precisar.	296,19	338,124	,132	,710
Recebi treinamento sobre como encontrar dados que eu possa potencialmente reutilizar.	296,09	329,007	,297	,703
Tenho acesso a treinamento formal sobre habilidades úteis para reutilizar dados.	296,00	329,652	,302	,703
É receptiva quanto ao reuso de dados.	294,71	337,917	,164	,709

Tem o reuso de dados como prática comum.	295,33	341,789	,049	,712
Convencionalmente reusa dados.	295,34	340,084	,100	,711
Não tem interesse em reusar dados.	296,37	344,179	-,018	,713
Incentivar-me a reusar dados.	293,66	343,852	,006	,712
Apoiar o reuso de dados.	293,54	342,745	,075	,711
Reusar dados frequentemente.	293,77	344,382	-,024	,713

FONTE: Dados da pesquisa com auxílio do *software* SPSS Statistics.

APÊNDICE 4 – ESTATÍSTICA DESCRITIVA

Parte I – Perfil dos pesquisadores, parte II comportamento sobre dados científicos de pesquisa e parte III comportamento dos pesquisadores sobre reuso de dados científicos de pesquisa.

Estatísticas de item

	Média	Erro Desvio	N
Qual é a sua principal ocupação?	2,10	,995	70
Qual é a sua região de trabalho?	2,36	1,551	70
Qual é a sua área de atuação?	3,37	2,600	70
Principal tipo de abordagem metodológica?	1,33	,847	70
Qual setor profissional está vinculado?	,47	,974	70
Qual é o principal financiador da sua pesquisa?	2,39	1,820	70
Qual é o tipo de dado que você mais utiliza?	3,00	2,455	70
Por que todos os seus dados não estão disponíveis?	6,89	2,262	70
Estou satisfeito com o meu processo de coleta de dados.	3,80	,773	70
Estou satisfeito com o processo para coletar os meus dados.	3,51	,794	70
Estou satisfeito com o processo para armazenar meus dados durante a vida do projeto (curto prazo).	3,63	,920	70
Estou satisfeito com o processo para armazenar meus dados durante a vida do projeto (curto prazo).	3,60	1,122	70
Estou satisfeito com o processo para análise dos meus dados.	3,83	,701	70

Estou satisfeito com o processo para análise dos meus dados.	3,59	1,097	70
Outros podem acessar facilmente meus dados.	3,20	1,058	70
Estou satisfeito com as ferramentas para preparar meus dados.	3,43	,894	70
Estou satisfeito com as ferramentas para preparar a documentação dos meus dados.	3,33	,928	70
Minha instituição ou grupo de pesquisa tem um processo formal estabelecido para gestão de dados durante a vida do projeto (curto prazo).	3,01	1,123	70
Minha instituição ou grupo de pesquisa tem um processo formal estabelecido para gestão de dados durante a vida do projeto (curto prazo).	2,73	1,115	70
Minha instituição ou grupo de pesquisa tem um processo formal estabelecido para armazenar os dados para além da vida do projeto (longo prazo).	3,77	1,416	70
Minha instituição ou grupo de pesquisa provê as ferramentas e suporte técnico para gestão de dados durante a vida do projeto (curto prazo).	3,31	1,110	70
Minha instituição ou grupo de pesquisa provê as ferramentas e suporte técnico para gestão de dados para além da vida da pesquisa (longo prazo).	2,93	1,133	70
Minha instituição ou grupo de pesquisa provê treinamentos e melhores práticas em gestão de dados.	2,46	1,059	70

Minha instituição ou grupo de pesquisa provê necessários para a gestão de dados durante a vida do projeto (curto prazo).	2,76	1,013	70
Minha instituição ou grupo de pesquisa provê recursos necessários para a gestão de dados para além da vida do projeto (longo prazo).	2,74	1,031	70
A falta de acesso aos dados gerados por outros pesquisadores ou instituições é o principal impedimento para o progresso da ciência.	3,14	1,026	70
A falta de acesso aos dados gerados por outros pesquisadores ou instituições restringe a minha habilidade de responder a questionamentos.	3,41	,985	70
Os dados podem ser mal interpretados devido a sua complexidade.	3,33	,880	70
Os dados podem ser mal interpretados devido a sua baixa qualidade.	3,69	,877	70
Os dados podem ser usados de maneiras adversas além das pretendidas.	3,33	,829	70
Eu usaria conjuntos de outros pesquisadores se eles fossem facilmente acessíveis.	4,57	,554	70
Eu estaria disposto a colocar pelo menos alguns dos meus dados em um repositório de dados sem restrições.	4,41	,925	70
Eu estaria disposto a colocar todos os meus dados em um repositório de dados sem restrições.	3,93	1,231	70

Estaria mais favorável a colocar meus dados disponíveis se eu pudesse acrescentar condições de acesso.	4,13	1,203	70
Estou satisfeito com minha habilidade em integrar dados de diferentes fontes para minhas questões de pesquisa.	3,66	,832	70
Estaria disposto a compartilhar dados em um amplo grupo de pesquisadores que utilizam dados de diferentes formas.	4,47	,583	70
É importante que meus dados sejam citados quando utilizados por outros pesquisadores.	4,86	,427	70
É apropriado criar novos conjuntos de dados a partir de dados compartilhados.	4,59	,625	70
Qual é o principal padrão de metadados utilizado para descrever seus dados?	1,97	2,648	70
Você é o único responsável por aprovar acesso aos seus dados?	1,17	,851	70
A principal agência de financiamento da sua pesquisa exige o fornecimento de um Plano de Gestão de Dados (PGD)?	,40	,788	70
Permite a verificação de resultados de pesquisa anteriores.	,74	,440	70
Contribui para o avanço científico.	,93	,259	70
Permite a reprodução de pesquisas.	,74	,440	70
Reduz os custos associados com a pesquisa.	,67	,473	70

Poupa o tempo do reutilizador.	,44	,500	70
Receber reconhecimento inferior.	1,44	,605	70
Ter minha esquisa mal considerada.	1,46	,674	70
Infringir involuntariamente os códigos éticos.	1,93	,890	70
Violar involuntariamente as normas de proteção de dados.	2,04	,939	70
Não ter consentimento informado pelos participantes de uma pesquisa.	2,30	1,012	70
Interpretar mal os dados.	2,81	,982	70
Fazer suposições ou associações incorretas com base nos dados.	2,97	,932	70
Usar dados indevidos.	2,44	,927	70
Ficar vulnerável a dados com erros ocultos.	3,50	,830	70
Ter capacidade limitada para identificar erros nos dados.	3,51	,864	70
Reutilizar dados imprecisos.	3,37	,981	70
Encontrar formas inovadores de abordar dados existentes.	3,26	1,031	70
Justificar a importância de um novo estudo com base nos dados existentes.	3,11	1,029	70
Identificar novos dados relevantes.	4,14	1,146	70
Selecionar dados reutilizáveis.	4,60	,788	70
Obter permissão para reutilizar dados.	3,40	,824	70
Recuperar dados e familiarizar-se com eles.	3,53	,793	70
Coletar e produzir dados por outras pessoas.	3,47	,847	70
Compreender o estudo original.	3,03	,900	70
Ajustar seu próprio projeto de pesquisa para acomodar os dados existentes.	3,50	,794	70

Alterar uma ideia de pesquisa para acomodar os dados existentes.	3,46	,811	70
Reorganizar seu estudo inicial para acomodar os dados existentes.	3,44	,862	70
a possibilidade de reprodução ou verificação de resultados.	,69	,468	70
tornar disponível de forma aberta o resultado de pesquisas realizadas com recursos públicos	,90	,302	70
permitir que outras pessoas façam novas perguntas baseadas nos dados existentes	,93	,259	70
avançar no estado de pesquisa e inovação	,74	,440	70
benefícios na carreira (por exemplo: reconhecimento e boa reputação)	,47	,503	70
aumento do número de citações	,60	,493	70
convites para colaboração com outros pesquisadores ou grupos de pesquisa	,54	,502	70
a possibilidade de transformação social na ciência cidadã	,41	,496	70
Ter documentação suficiente (por exemplo: livro de códigos).	4,56	,715	70
Ter documentação simples e clara.	4,74	,440	70
Estar na mesma unidade de análise que eu preciso (por exemplo: indivíduos, grupos).	4,19	1,011	70
Encaixar no meu estudo.	4,30	,983	70
Estar conforme o formato que eu preciso.	3,86	1,081	70
Ser originário de uma fonte confiável.	4,67	,531	70

Ser produto de uma fonte confiável.	4,13	,850	70
Ser consistente/estável.	4,23	,820	70
Estar completo ou com poucos dados ausentes.	4,71	,640	70
Ser produto de um estudo bem projetado	3,60	,875	70
Pertencer a um estudo que realizou um plano de pesquisa original.	2,71	1,092	70
Ser originado de um estudo bem executável.	3,44	,895	70
Estou ciente dos repositórios de onde posso reutilizar dados.	3,50	,897	70
Posso encontrar facilmente repositórios com dados relacionados à minha pesquisa.	3,10	,801	70
Os pesquisadores primários estão disponíveis para fornecer informações adicionais aos seus dados.	2,30	,906	70
Posso facilmente entrar em contato com os pesquisadores primários para obter esclarecimentos sobre seus dados, se eu precisar.	2,13	1,034	70
Recebi treinamento sobre como encontrar dados que eu possa potencialmente reutilizar.	2,23	1,265	70
Tenho acesso a treinamento formal sobre habilidades úteis para reutilizar dados.	2,31	1,198	70
É receptiva quanto ao reuso de dados.	3,60	,907	70
Tem o reuso de dados como prática comum.	2,99	,893	70
Convencionalmente reusa dados.	2,97	,900	70

Não tem interesse em reusar dados.	1,94	,720	70
Incentivar-me a reusar dados.	4,66	,508	70
Apoiar o reuso de dados.	4,77	,456	70
Reusar dados frequentemente.	4,54	,652	70